P450 OXYGENASES AND METHODS OF USE

REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No. 60/489,597, filed July 22, 2003, which application is incorporated herein in its entirety.

STATEMENT OF GOVERNMENT SUPPORT

This invention was made with United States government support pursuant to grant no. CA 55254, from the National Institutes of Health. The United States government has certain rights in the invention.

FIELD OF THE DISCLOSURE

The present disclosure relates to P450 oxygenases, particularly to taxoid 5α-hydroxylases, and the nucleic acids that encode them, and to methods of using such oxygenase nucleic acids and enzymes, for example, to produce Taxol™ (commonly known as paclitaxel) and other paclitaxel intermediates.

BACKGROUND OF THE DISCLOSURE

The complex diterpenoid Taxol™ (Bristol-Myers Squibb; common name paclitaxel) (Wani *et al., J. Am. Chem. Soc.* 93:2325-2327, 1971) is a potent antimitotic agent with excellent activity against a wide range of cancers, including ovarian and breast cancer (Arbuck and Blaylock, *Taxol: Science and Applications*, CRC Press, Boca Raton, 397-415, 1995; Holmes *et al., ACS Symposium Series* 583:31-57, 1995). Paclitaxel was isolated originally from the bark of the Pacific yew (*Taxus brevifolia*). For a number of years, paclitaxel was obtained exclusively from yew bark, but low yields of this compound from the natural source coupled to the destructive nature of the harvest, prompted new methods of paclitaxel production to be developed.

Total chemical syntheses of paclitaxel have been achieved (for review, see, Kingston *et al., Prog. Chem. Org. Nat. Prod.* 84:56-225, 2002) but the yields of the drug by this method are too low to be practical. Paclitaxel currently is produced primarily by chemical semisynthesis from advanced taxane metabolites (Holton *et al., Taxol: Science and Applications*, CRC Press, Boca Raton, 97-121, 1995; Hezari and Croteau, *Planta Medica*, 63:291-295, 1997) that are isolated from the needles (a renewable resource) of various *Taxus* speciea. However, at least because of the increasing demand for this drug both for use earlier in the course of cancer intervention and for new therapeutic applications (Goldspiel, *Pharmacotherapy* 17:110S-125S, 1997), high-yield, cost-effective methods of paclitaxel production continue to be needed. Some have proposed isolating paclitaxel from alternative biological sources, such as the endophytic fungi, *Taxomyces andreanae* (Stierle *et al., J. Nat. Prod.* 58:1315-1324, 1995), or from *Taxus* cell cultures (Ketchum *et al., Biotechnol. Bioeng.* 62:97-105, 1999). However, these methods are also too inefficient to produce sufficient quantities of the drug and have had limited commercial success.

Improving the production yield of paclitaxel from any biological system, whether intact organisms (such as, *Taxus* plants or paclitaxel-producing fungi) or cell cultures, would be facilitated by a detailed understanding of the paclitaxel biosynthetic pathway. The paclitaxel biosynthetic pathway is complex and believed to involve nearly 20 distinct steps (Floss and Mocek, *Taxol: Science and Applications*, CRC Press, Boca Raton, 191-208, 1995; and Croteau *et al.*, *Curr. Top. Plant Physiol.* 15:94-104, 1996). However, relatively few of the enzymatic reactions and intermediates of this complicated pathway have been defined in detail.

The first committed enzyme of the paclitaxel pathway is believed to be taxadiene synthase (Koepp *et al.*, *J. Biol. Chem.* 270:8686-8690, 1995), which cyclizes the common precursor geranylgeranyl diphosphate (Hefner *et al.*, *Arch. Biochem. Biophys.* 360:62-74, 1998) to taxadiene (FIG. 1). The cyclized intermediate (*i.e.*, taxa-4(5),11(12)-diene) subsequently undergoes modification involving at least eight oxygenation steps, a formal dehydrogenation, an epoxide rearrangement to an oxetane, and several acylations (Floss and Mocek, *Taxol: Science and Applications*, CRC Press, Boca Raton, 191-208, 1995; and Croteau *et al.*, *Curr. Top. Plant Physiol.* 15:94-104, 1996). Taxadiene synthase has been isolated from *T. brevifolia* and characterized (Hezari *et al.*, *Arch. Biochem. Biophys.* 322:437-444, 1995), the mechanism of action defined (Lin *et al.*, *Biochemistry* 35:2968-2977, 1996), and the corresponding cDNA clone isolated and expressed (Wildung and Croteau, *J. Biol. Chem.* 271:9201-9204, 1996).

The second specific step of paclitaxel biosynthesis is believed to be an oxygenation (hydroxylation) reaction that introduces a hydroxyl group to position 5 of taxa-4(5),11(12)-diene to produce taxa-4(20),11(12)-dien-5α-ol. Using a crude *Taxus* microsome preparation, Hefner *et al.* (*Methods Enzymol.* 272:243-250, 1996) demonstrated a microsomal activity that catalyzed the stereospecific hydroxylation of taxa-4(5),11(12)-diene to taxa-4(20),11(12)-dien-5α-ol (with double-bond rearrangement) (Hefner *et al.*, *Chem. Biol.*, 3:479-489, 1996). This microsomal activity was attributed to one or more cytochrome P450 oxygenases (Hefner *et al.*, *Chemistry and Biology* 3:479-489, 1996). Cytochrome P450 oxygenases are enzymes that have a unique sulfur atom ligated to the heme iron and that, when reduced, form carbon monoxide (CO) complexes. When complexed to carbon monoxide, cytochrome P450 proteins display a major absorption peak (Soret band) near 450 nm.

Taxus microsomal preparations were further shown to catalyze the hydroxylation of taxadiene or taxadien-5α-ol to the level of a pentaol (Hefner *et al.*, *Methods Enzymol.* 272:243-250, 1996; Lovy Wheeler *et al.*, *Arch. Biochem. Biophys.*, 390:265-278, 2001). These results suggested that the paclitaxel biosynthetic pathway included at least five distinct cytochrome P450 taxoid oxygenases in the early parts of the pathway (Hezari *et al.*, *Planta Med.* 63:291-295, 1997). Later steps of the paclitaxel biosynthetic pathway are thought to include at least three additional oxygenation steps (C1 and C7 hydroxylations and an epoxidation at C4-C20). These steps also are believed to be catalyzed by cytochrome P450 enzymes, but these reactions reside too far down the pathway to observe in microsomes by current experimental methods (Croteau *et al.*, *Curr. Topics*

Plant Physiol. 15:94-104, 1995; Hezari *et al.*, *Planta Med.* 63:291-295, 1997 Lovy Wheeler *et al.*, *Arch. Biochem. Biophys.*, 390:265-278, 2001). Since *Taxus* (yew) plants and cells do not appear to accumulate taxoid metabolites bearing fewer than six oxygen atoms (*e.g.*, hexaol or epoxypentaol) (Kingston *et al.*, *Prog. Chem. Org. Nat. Prod.* 61:1-206, 1993), such intermediates must be rapidly transformed down the pathway, indicating that the oxygenations (hydroxylations) are relatively slow pathway steps.

Taxus microsome preparations contain hundreds of different proteins, including an estimated 30 to 50 similar cytochrome P450 oxygenases (Hefner *et al.*, *Methods Enzymol.* 272:243-250, 1996). Biochemical purification of cytochrome P450 enzymes from *Taxus* microsomes (Hefner *et al.*, *Methods Enzymol.* 272:243-250, 1996) is not practical, at least, because the numerous P450 cytochrome oxygenases present in this cell fraction have very similar physical properties (Mihaliak *et al.*, *Methods Plant Biochem.* 9:261-279, 1993). With no useful biochemical means to distinguish among the many microsomal P450 oxygenases, it is not feasible to sufficiently purify any one enzyme to obtain even short peptide sequences. As a result, other methods are needed to isolate and characterize these important enzymes at the molecular level.

Differential display reverse transcription PCR (DD-RT PCR) has been used to isolate methyl jasmonate-induced nucleic acids encoding taxoid oxygenases of the paclitaxel biosynthetic pathway (see, for example, PCT Pub. No. WO01/34780). Several of the encoded oxygenase enzymes have been expressed and functionally characterized (PCT Pub. No. WO01/34780; Schoendorf *et al.*, *Proc. Natl. Acad. Sci. USA*, 98:1501-1506, 2001; Jennewein *et al.*, *Proc. Natl. Acad. Sci. USA*, 98:13595-13600, 2001; Jennewein *et al.*, Arch. Biochem. Biophys., 413:262-270, 2003). However, transcripts encoding taxoid oxygenases that are not, or only weakly, induced are likely to be missed by the DD-RT PCR technique.

Paclitaxel is an important drug that is not efficiently produced using current methods. Genetic engineering and recombinant technologies offer ways to increase paclitaxel and taxoid yields. To capitalize on these technologies, there is a continuing need to identify and isolate the genes encoding the enzymes of the paclitaxel biosynthetic pathway, including, for example, the numerous oxygenase enzymes, and for methods of using such genes and enzymes to produce paclitaxel and its intermediates.

SUMMARY OF THE DISCLOSURE

This disclosure provides a novel P450 oxygenase, which is capable of incorporating oxygen (for example, a hydroxyl group or epoxide ring) into a substrate, such as a taxoid. In some examples, the disclosed oxygenase incorporates oxygen at the C5 position of a taxoid, wherein the disclosed oxygenase is referred to as a taxoid 5-hydroxylase. In more specific examples, the oxygen is incorporated in the alpha configuration at C5 of a taxoid, wherein the disclosed oxygenase is referred to as a taxoid 5α-hydroxylase. In some examples, a taxoid substrate for a disclosed oxygenase includes a taxadiene, such as taxa-4(5),11(12)-diene and taxa-4(20),11(12)-diene.

Encompassed within this disclosure are the protein and nucleic acid sequences of the disclosed P450 oxygenase. Also provided are nucleotide and amino acid sequence variants, oligonucleotides and protein fragments. This disclosure demonstrates that the disclosed enzymes catalyze the oxygenation of taxoids, for example, at the C5 position. Evidence provided herein also demonstrates the relaxed substrate specificity of the disclosed oxygenases. It is also disclosed that a nucleic acid encoding a disclosed oxygenase, such as a 5α-hydroxylase cDNA, can be operatively linked to a promoter and cells can be transfected with the recombinant polynucleotide.

Also provided herein are methods for using a disclosed oxygenase, such as a taxoid 5α-hydroxylase. Such methods include, without limitation, methods of using a taxoid 5α-hydroxylase to hydroxylate a taxoid substrate or to produce (or increase the yield of) paclitaxel or paclitaxel intermediates. Examples of these methods include introducing a taxoid 5α-hydroxylase recombinant polynucleotide into a cell, such as a *Taxus* cell, or contacting a taxoid with a 5α-hydroxylase polypeptide or functional fragment thereof. Also disclosed are taxoid 5α-hydroxylase-specific binding agents.

The foregoing and other features and advantages will become more apparent from the following detailed description of several embodiments, which proceeds with reference to the accompanying figures.

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1 shows an outline of the early steps of paclitaxel biosynthesis. Paclitaxel (1) formation involves the cyclization of geranylgeranyl diphosphate (2) to taxa-4(5),11(12)-diene (3) and cytochrome P450-mediated hydroxylation to taxa-4(20),11(12)-dien-5α-ol (4). The multiple arrows between taxa-4(20),11(12)-dien-5α-ol (4) and paclitaxel (1) are represented of numerous additional enzymatic steps in the paclitaxel biosynthetic pathway.

FIG. 2 shows an alignment of the deduced amino acid sequences of selected taxoid hydroxylases. The sequences of taxoid 10β-hydroxylase (T10H; SEQ ID NO: 16), taxoid 13α-hydroxylase (T13H; SEQ ID NO: 18) and the clone S1 taxadiene 5α-hydroxylase (T5H; SEQ ID NO: 2) are compared. Black boxes indicate identical residues for the three sequences; grey boxes indicate identical residues for two of the three.

FIG. 3 shows substrate binding spectra of taxadiene isomers. Microsomes from *S. frugiperda* cells enriched with 200 pmol recombinant taxoid 5α-hydroxylase (clone S1) were employed. Taxa-4(20),11(12)-diene was assayed over a concentration range from 0.1 to 10 μM, and a binding constant (Ks) of 4 ± 1 μM was determined (upper graph). Taxa-4(5),11(12)-diene was assayed over a concentration range of 0.1 to 20 μM and a binding constant (Ks) of 6.5 ± 1.5 μM was determined (lower graph).

FIG. 4 shows a kinetic evaluation of taxadiene isomers. Taxa-4(20),11(12)-diene (○) and taxa-4(5),11(12)-diene (●) were evaluated with microsomes from *S. frugiperda* cells enriched with 50 pmol recombinant taxoid 5α-hydroxylase (A), and with microsomes from *T. media* suspension

cells containing about 50 pmol of total native cytochrome P450 (B). Substrate concentration range was varied from 1 to 500 μM in all cases. Taxa-4(20),11(12)-diene yielded an average Km value of 21.5 μM with Vrel of 135, and taxa-4(5),11,12-diene yielded an average Km value of 36 μM with Vrel of 100.

FIG. 5 shows a proposed, but not binding, mechanism for cytochrome P450 taxoid 5α-hydroxylase. This cytochrome P450-mediated conversion of taxa-4(5),11(12)-diene (3) and taxa-4(20),11(12)-diene (5) to taxa-4(20),11(12)-dien-5α-ol (4) is believed to involve hydrogen abstraction from C20 (in 3) or C5 (in 5) to provide a common allylic radical intermediate, followed by oxygen insertion at the 5α-face to yield taxadien-5α-ol (4). Isomerization of 3 to 5 was not observed, nor does the route via epoxide 6 with rearrangement seem likely.

FIGS. 6A and 6B collectively show an alignment of the deduced amino acid sequences of selected taxoid oxygenases. The sequence of clone S1 taxoid 5α-hydroxylase (T5H; SEQ ID NO: 2) is compared to the sequences of eight other taxoid oxygenases isolated from *Taxus cuspidata*. Each of the illustrated oxygenases is known to have a positive CO difference spectrum and to oxidize intermediates in the paclitaxel biosynthetic pathway or derivatives thereof (see, *e.g.*, PCT Pub. No. WO01/23586, Table 2). Oxygenase sequences other than T5H are designated as in PCT Pub. No. WO01/23586 (herein, "F clones"). F31 is a taxoid 7β-hydroxylase (SEQ ID NO: 8); F72 is a taxoid 14β-hydroxylase (SEQ ID NO: 12), F14 is a taxoid 10β-hydroxylase (SEQ ID NO: 16), and F16 is a taxoid 13α-hydroxylase (SEQ ID NO: 18). F14 and F16 correspond to T10H and T13H, respectively, in FIG. 2. The other illustrated F clones are F12 (SEQ ID NO: 4), F21 (SEQ ID NO: 6), F51 (SEQ ID NO: 10), and F9 (SEQ ID NO: 14). Shaded amino acid residues are identical among all sequences; ":" indicates conservative substitutions among the amino acid residues at that position.

BRIEF DESCRIPTION OF THE SEQUENCE LISTING

The nucleic and amino acid sequences listed in the accompanying sequence listing are shown using standard letter abbreviations for nucleotide bases, and three letter code for amino acids, as defined in 37 C.F.R. §1.822. Only one strand of each nucleic acid sequence is shown, but the complementary strand is understood as included by any reference to the displayed strand. In the accompanying sequence listing:

SEQ ID NO: 1 shows a nucleic acid sequence encoding a taxoid 5α-hydroxylase (GenBank Accession No. AY289209) and its corresponding amino acid sequence.

SEQ ID NO: 2 shows an amino acid sequence of a taxoid 5α-hydroxylase (GenBank Accession No. AAQ56240.1), which is encoded by the nucleic acid sequence in SEQ ID NO: 1.

SEQ ID NO: 3 shows the nucleic acid sequence of Clone F12 described in PCT Pub. No. WO01/23586.

SEQ ID NO: 4 shows the taxoid oxygenase amino acid sequence encoded by Clone F12.

SEQ ID NO: 5 shows the nucleic acid sequence of Clone F21 described in PCT Pub. No. WO01/23586.

SEQ ID NO: 6 shows the taxoid oxygenase amino acid sequence encoded by Clone F21.

SEQ ID NO: 7 shows a nucleic acid sequence encoding a taxoid 7β-hydroxylase (see, also, clone F31 as described in PCT Pub. No. WO01/23586) (GenBank Accession No. AY307951).

SEQ ID NO: 8 shows a taxoid 7β-hydroxylase amino acid sequence (GenBank Accession No. AAQ7553), which is encoded by the nucleic acid sequence in SEQ ID NO: 7.

SEQ ID NO: 9 shows the nucleic acid sequence of Clone F51 described in PCT Pub. No. WO01/23586.

SEQ ID NO: 10 shows the taxoid oxygenase amino acid sequence encoded by Clone F51.

SEQ ID NO: 11 shows a nucleic acid sequence encoding a taxoid 14β-hydroxylase (see, also, clone F72 as described in PCT Pub. No. WO01/23586) (GenBank Accession No. AY188177; Jennewein *et al.*, *Arch. Biochem. Biophys.*, 413(2):262-270, 2003).

SEQ ID NO: 12 shows a taxoid 14β-hydroxylase amino acid sequence (GenBank Accession No. AAO66199), which is encoded by the nucleic acid sequence in SEQ ID NO: 11.

SEQ ID NO: 13 shows the nucleic acid sequence of Clone F9 described in PCT Pub. No. WO01/23586.

SEQ ID NO: 14 shows the taxoid oxygenase amino acid sequence encoded by Clone F9.

SEQ ID NO: 15 shows a nucleic acid sequence encoding a taxoid 10β-hydroxylase (see, also, clone F14 as described in PCT Pub. No. WO01/23586) (GenBank Accession No. AY563635; Jennewein *et al.*, *Proc. Natl. Acad. Sci. USA*, 101(24):9149-9154, 2004).

SEQ ID NO: 16 shows a taxoid 10β-hydroxylase amino acid sequence (GenBank Accession No. AAT47183) encoded by the nucleic acid sequence in SEQ ID NO: 15.

SEQ ID NO: 17 shows a nucleic acid sequence encoding a taxoid 13α-hydroxylase (see, also, clone F16 as described in PCT Pub. No. WO01/23586) (GenBank Accession No. AY056019; Jennewein *et al.*, *Proc. Natl. Acad. Sci. USA*, 98(24):13595-13600, 2001).

SEQ ID NO: 18 shows a taxoid 13α-hydroxylase amino acid sequence (GenBank Accession No. AAL23619) encoded by the nucleic acid sequence in SEQ ID NO: 17.

SEQ ID NO: 19 shows a nucleic acid sequence encoding a taxadiene synthase (GenBank Accession No. U48796; Wildung and Croteau, *J. Biol. Chem.*, 271(16):9201-9204, 1996) and its corresponding amino acid sequence.

SEQ ID NO: 20 shows the amino acid sequence of a taxadiene synthase (GenBank Accession No. AAC49310), which is encoded by the nucleic acid sequence in SEQ ID NO: 19.

SEQ ID NO: 21 shows a nucleic acid sequence encoding a taxadienol acetyl transferase (also called, TAT or TAX1) (GenBank Accession No. AF190130; Walker *et al.*, *Arch. Biochem. Biophys.*, 374(2):371-380, 2000) and its corresponding amino acid sequence.

SEQ ID NO: 22 shows the amino acid sequence of a taxadienol acetyl transferase (GenBank Accession No. AAF34254), which is encoded by the nucleic acid sequence in SEQ ID NO: 21.

SEQ ID NO: 23 shows a nucleic acid sequence encoding a 2-debenzoyl-7,13-diacetylbaccatin III-2-O-benzoyl transferase (also called, TAX2) (GenBank Accession

No. AF297618; Walker and Croteau, *Proc. Natl. Acad. Sci. USA*, 97(25):13591-13596, 2000)) and its corresponding amino acid sequence.

SEQ ID NO: 24 shows the amino acid sequence of a 2-debenzoyl-7,13-diacetylbaccatin III-2-O-benzoyl transferase (GenBank Accession No. AAG38049), which is encoded by the nucleic acid sequence in SEQ ID NO: 23.

SEQ ID NOs: 25-29 show primers directed to the commonly occurring P450 oxygenase PERF motif and its variant forms.

SEQ ID NOs: 30-31 show primers directed to the conserved P450 oxygenase heme-binding region.

SEQ ID NOs: 32-33 show primers suitable for amplifying a nucleic acid sequence encoding a taxoid 5α-hydroxylase.

SEQ ID NO: 34 shows a nucleic acid sequence encoding a 10-deacetylbaccatin III-10-O-acetyl transferase (also called, TAX6 or DBAT) (GenBank Accession No. AF193765; *e.g.*, Walker and Croteau, *Proc. Natl. Acad. Sci. USA*, 97(2):583-587, 2000) and its corresponding amino acid sequence.

SEQ ID NO: 35 shows the amino acid sequence of a 10-deacetylbaccatin III-10-O-acetyl transferase, which is encoded by the nucleic acid sequence in SEQ ID NO: 33.

SEQ ID NO: 36 shows a nucleic acid sequence encoding a taxoid 13-phenylpropanoyltransferase (also called, TAX7) (GenBank Accession No. AY082804, Walker *et al.*, *Proc. Natl. Acad. Sci. USA*, 99(20):12715-12720, 2002) and its corresponding amino acid sequence.

SEQ ID NO: 37 shows the amino acid sequence of a taxoid 13-phenylpropanoyltransferase, which is encoded by the nucleic acid sequence in SEQ ID NO: 35.

SEQ ID NO: 38 shows a nucleic acid sequence encoding a taxoid 3'-N-debenzoyltaxol N-benzoyltransferase (also called, TAX10 or DBNTBT) (GenBank Accession No. AF466397; Walker *et al.*, *Proc. Natl. Acad. Sci. USA*, 99(14):9166-9171, 2002) and its corresponding amino acid sequence.

SEQ ID NO: 39 shows the amino acid sequence of a taxoid 3'-N-debenzoyltaxol N-benzoyltransferase, which is encoded by the nucleic acid sequence in SEQ ID NO: 37.

SEQ ID NO: 40 shows a nucleic acid sequence encoding a taxoid 2α-hydroxylase ((GenBank Accession No. AY518383; Chau and Croteau, *Arch. Biochem. Biophys.*, 427(1):48-57, 2004) and its corresponding amino acid sequence.

SEQ ID NO: 41 shows the amino acid sequence of a taxoid 2α-hydroxylase, which is encoded by the nucleic acid sequence in SEQ ID NO: 39.

DETAILED DESCRIPTION

I. Abbreviations and Terms

CO	carbon monoxide
ELISA	enzyme-linked immunosorbent assay

GC-MS	gas chromatography-mass spectroscopy
HPLC	high performance liquid chromatography
HSQC	heteronuclear single quantum coherence
KIE	kinetic isotope effect
kDa	kilodaltons
MW	molecular weight
NMR	nuclear magnetic reasonance spectroscopy
ORF	open reading frame
RACE	rapid analysis of cDNA ends
ROESY	rotational nuclear overhauser effect spectroscopy
TLC	thin layer chromatography
TOCSY	total correlated spectroscopy

Unless otherwise noted, technical terms are used according to conventional usage. Definitions of common terms in molecular biology may be found in Benjamin Lewin, *Genes V*, published by Oxford University Press, 1994 (ISBN 0-19-854287-9); Kendrew *et al.* (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); and Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

In order to facilitate review of the various embodiments disclosed herein, the following explanations of specific terms are provided:

Amplification: When used in reference to nucleic acids, techniques that increase the number of copies of a nucleic acid molecule in a sample or specimen. An example of amplification is the polymerase chain reaction, in which a biological sample collected from a subject is contacted with a pair of oligonucleotide primers, under conditions that allow for the hybridization of the primers to nucleic acid template in the sample. The primers are extended under suitable conditions, dissociated from the template, and then re-annealed, extended, and dissociated to amplify the number of copies of the nucleic acid. The product of *in vitro* amplification can be characterized by electrophoresis, restriction endonuclease cleavage patterns, oligonucleotide hybridization or ligation, and/or nucleic acid sequencing, using standard techniques. Other examples of *in vitro* amplification techniques include strand displacement amplification (see U.S. Patent No. 5,744,311); transcription-free isothermal amplification (see U.S. Patent No. 6,033,881); repair chain reaction amplification (see WO 90/01069); ligase chain reaction amplification (see EP-A-320 308); gap filling ligase chain reaction amplification (see U.S. Patent No. 5,427,930); coupled ligase detection and PCR (see U.S. Patent No. 6,027,889); and NASBA™ RNA transcription-free amplification (see U.S. Patent No. 6,025,134).

Binding or stable binding: An oligonucleotide binds or stably binds to a target nucleic acid if a sufficient amount of the oligonucleotide forms base pairs or is hybridized to its target nucleic acid, to permit detection of that binding. Binding can be detected by either physical or functional properties of the target:oligonucleotide complex. Binding between a target and an oligonucleotide

can be detected by any procedure known to one of ordinary skill in the art, including both functional and physical binding assays. Binding can be detected functionally by determining whether binding has an observable effect upon a biosynthetic process such as expression of a gene, DNA replication, transcription, translation and the like.

Physical methods of detecting the binding of complementary strands of DNA or RNA are well known in the art, and include such methods as DNase I or chemical footprinting, gel shift and affinity cleavage assays, Northern blotting, dot blotting and light absorption detection procedures. For example, one method that is widely used, because it is so simple and reliable, involves observing a change in light absorption of a solution containing an oligonucleotide (or an analog) and a target nucleic acid at 220 to 300 nm as the temperature is slowly increased. If the oligonucleotide or analog has bound to its target, there is a sudden increase in absorption at a characteristic temperature as the oligonucleotide (or analog) and the target disassociate from each other, or melt.

The binding between an oligomer and its target nucleic acid is frequently characterized by the temperature (T_m) at which 50% of the oligomer is melted from its target. A higher T_m means a stronger or more stable complex relative to a complex with a lower T_m.

cDNA (complementary DNA): A piece of DNA lacking internal, non-coding segments (introns) and transcriptional regulatory sequences. cDNA can also contain untranslated regions (UTRs) that can be responsible for translational control in the corresponding RNA molecule. cDNA can be synthesized in the laboratory by reverse transcription from messenger RNA extracted from cells.

DNA (deoxyribonucleic acid): A long chain polymer which comprises the genetic material of most living organisms (some viruses have genes comprising ribonucleic acid (RNA)). The repeating units in DNA polymers are four different nucleotides, each of which comprises one of the four bases, adenine, guanine, cytosine and thymine bound to a deoxyribose sugar to which a phosphate group is attached. Triplets of nucleotides (referred to as codons) code for each amino acid in a polypeptide. The term codon is also used for the corresponding (and complementary) sequences of three nucleotides in the mRNA into which the DNA sequence is transcribed.

Unless otherwise specified, any reference to a DNA molecule is intended to include the reverse complement of that DNA molecule. Except where single strandedness is required by the text herein, DNA molecules, though written to depict only a single strand, encompass both strands of a double-stranded DNA molecule. Thus, a reference to the nucleic acid molecule that encodes a specific protein, or a fragment thereof, encompasses both the sense strand and its reverse complement. Thus, for instance, it is appropriate to generate probes or primers from the reverse complement sequence of the disclosed nucleic acid molecules.

Encode: A polynucleotide is said to "encode" a polypeptide if, in its native state or when manipulated by methods well known to those of ordinary skill in the art, it can be transcribed and/or translated to produce the mRNA for and/or the polypeptide or a fragment thereof. The anti-sense strand is the complement of such a nucleic acid, and the encoding sequence can be deduced therefrom.

Functional fragments and variants of a polypeptide: Included are those fragments and variants that maintain one or more functions of the parent polypeptide. It is recognized that the gene or cDNA encoding a polypeptide can be considerably mutated without materially altering one or more the polypeptide's functions. First, the genetic code is well-known to be degenerate, and thus different codons encode the same amino acids. Second, even where an amino acid substitution is introduced, the mutation can be conservative and have no material impact on the essential functions of a protein (see, Stryer, *Biochemistry*, Third Edition, W.H. Freeman and Company, New York, N.Y., p. 769, 1988). Third, part of a polypeptide chain can be deleted without impairing or eliminating all of its functions. Fourth, insertions or additions can be made in the polypeptide chain for example, adding epitope tags, without impairing or eliminating its functions (Ausubel *et al.*, *Current Protocols in Molecular Biology*, Greene Publ. Assoc. and Wiley-Intersciences, 1997; Jennewein *et al.*, *Arch. Biochem. Biophys.*, 413:262-270, 2003). Other modifications that can be made without materially impairing one or more functions of a polypeptide include, for example, *in vivo* or *in vitro* chemical and biochemical modifications or the incorporation of unusual amino acids. Such modifications include, for example, acetylation, carboxylation, phosphorylation, glycosylation, ubiquination, labeling, *e.g.*, with radionucleides, and various enzymatic modifications, as will be readily appreciated by ordinarily skilled artisans. A variety of methods for labeling polypeptides and labels, which are useful for such purposes are well known in the art, and include radioactive isotopes such as ^{32}P, ligands that bind to or are bound by labeled specific binding partners (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands. Functional fragments and variants can be of varying length. For example, some functional fragments have at least 75,100, 200, 300 or 400 amino acid residues.

A functional fragment or variant of a disclosed P450 oxygenase, such as a taxoid 5α-hydroxylase, is defined herein as a polypeptide capable of oxidizing (for example, hydroxylating or epoxidizing) a taxoid. In specific examples, a functional fragment or variant oxidizes a taxoid at the C5 position. It includes any polypeptide of about 100 or more amino acid residues in length, which is capable of having taxoid oxygenase activity.

Heterologous: A type of sequence that is not normally (*i.e.* in the wild-type sequence) found adjacent to a second sequence. In one embodiment, the sequence is from a different genetic source, such as a virus or organism, than the second sequence.

Host cell: Any cell that is capable of being transformed with a recombinant nucleic acid sequence. For example, bacterial cells, fungal cells, plant cells (such as, *Taxus* cells), insect cells, avian cells, mammalian cells, and amphibian cells. A host cell can be isolated or it can exist as a part of a transgenic organism (such as, microorganism (or lower life form) or a macroorganism). In specific examples, a host cell can be a primary cell or a cell line. A primary cell is a cell that is taken directly from a living organism, such as a plant (*e.g.*, a plant from the genus *Taxus*), which is not immortalized. The term "cell line" refers to a cell that is able to replicate in culture. Some cell lines (often called immortal cells) are capable of an essentially unlimited number of cell divisions. A primary cell may become a cell line upon continuous culture. In some instances, an immortal cell can

arise spontaneously, for example, as a result of uncharacterized alterations in the cell genome. In other case, a cell, such as a primary cell, can be made immortal using techniques commonly known in the art, including transfection with SV40 T-antigen or telomerase reverse transcriptase (TERT) (for review, *e.g.*, Hahn, *Mol. Cells*, 13(3):351-361, 2002).

Hybridization: Oligonucleotides and other nucleic acids hybridize by hydrogen bonding, which includes Watson-Crick, Hoogsteen or reversed Hoogsteen hydrogen bonding, between complementary bases. Generally, nucleic acid consists of nitrogenous bases that are either pyrimidines (cytosine (C), uracil (U), and thymine (T)) or purines (adenine (A) and guanine (G)). These nitrogenous bases form hydrogen bonds between a pyrimidine and a purine, and the bonding of the pyrimidine to the purine is referred to as **base pairing.** More specifically, A will hydrogen bond to T or U, and G will bond to C. **Complementary** refers to the base pairing that occurs between to distinct nucleic acid sequences or two distinct regions of the same nucleic acid sequence.

Specifically hybridizable and **specifically complementary** are terms that indicate a sufficient degree of complementarity such that stable and specific binding occurs between a first nucleic acid (such as, an oligonucleotide) and a DNA or RNA target. The first nucleic acid (such as, an oligonucleotide) need not be 100% complementary to its target sequence to be specifically hybridizable. A first nucleic acid (such as, an oligonucleotide) is specifically hybridizable when there is a sufficient degree of complementarity to avoid non-specific binding of the first nucleic acid (such as, an oligonucleotide) to non-target sequences under conditions where specific binding is desired. Such binding is referred to as specific hybridization.

Hybridization conditions resulting in particular degrees of stringency will vary depending upon the nature of the hybridization method of choice and the composition and length of the hybridizing nucleic acid sequences. Generally, the temperature of hybridization and the ionic strength (especially the Na^+ concentration) of the hybridization buffer will determine the stringency of hybridization, though waste times also influence stringency. Calculations regarding hybridization conditions required for attaining particular degrees of stringency are discussed by Sambrook *et al.* (ed.), *Molecular Cloning: A Laboratory Manual*, 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989, chapters 9 and 11.

The following are exemplary sets of hybridization conditions and are not meant to be limiting.

Very High Stringency (detects sequences that share 90% sequence identity)

Hybridization:	5x SSC at 65°C for 16 hours
Wash twice:	2x SSC at room temperature (RT) for 15 minutes each
Wash twice:	0.5x SSC at 65°C for 20 minutes each

High Stringency (detects sequences that share 80% sequence identity or greater)

Hybridization:	5x-6x SSC at 65°C-70°C for 16-20 hours
Wash twice:	2x SSC at RT for 5-20 minutes each
Wash twice:	1x SSC at 55°C-70°C for 30 minutes each

Low Stringency (detects sequences that share greater than 50% sequence identity)

Hybridization:	6x SSC at RT to 55°C for 16-20 hours
Wash at least twice:	2x-3x SSC at RT to 55°C for 20-30 minutes each.

Isolated: A biological component (such as a nucleic acid molecule, protein or organelle) that has been substantially separated or purified away from other biological components in the cell of the organism in which the component naturally occurs, *i.e.*, other chromosomal and extra-chromosomal DNA and RNA, proteins and organelles. Nucleic acids and proteins that have been isolated include nucleic acids and proteins purified by standard purification methods. The term also embraces nucleic acids and proteins prepared by recombinant expression in a host cell as well as chemically synthesized nucleic acids.

Nucleotide: This term includes, but is not limited to, a monomer that includes a base linked to a sugar, such as a pyrimidine, purine or synthetic analogs thereof, or a base linked to an amino acid, as in a peptide nucleic acid (PNA). A nucleotide is one monomer in a polynucleotide. A nucleotide sequence refers to the sequence of bases in a polynucleotide.

Oligonucleotide: A plurality of joined nucleotides joined by native phosphodiester bonds, between about 6 and about 300 nucleotides in length. An oligonucleotide analog refers to moieties that function similarly to oligonucleotides but have non-naturally occurring portions. For example, oligonucleotide analogs can contain non-naturally occurring portions, such as altered sugar moieties or inter-sugar linkages, such as a phosphorothioate oligodeoxynucleotide. Functional analogs of naturally occurring polynucleotides can bind to RNA or DNA, and include peptide nucleic acid (PNA) molecules.

Particular oligonucleotides and oligonucleotide analogs can include linear sequences up to about 200 nucleotides in length, for example a sequence (such as DNA or RNA) that is at least 6 bases, for example at least 8, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100 or even 200 bases long, or from about 6 to about 50 bases, for example about 10-25 bases, such as 12, 15 or 20 bases.

Operably linked: A first nucleic acid sequence is operably linked with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two protein-coding regions, in the same reading frame.

Open reading frame (ORF): A series of nucleotide triplets (codons) coding for amino acids without any internal termination codons. These sequences are usually translatable into a peptide.

Organism: Any individual living thing, whether unicellular or multi-cellular and including all members of Archaea, Bacteria, and Eukaryota taxonomical classifications, such as plants, yeast, bacteria, fungi, and insects.

Ortholog: Two nucleic acid or amino acid sequences are orthologs of each other if they share a common ancestral sequence and diverged when a species carrying that ancestral sequence split into two species. Orthologous sequences are also homologous sequences.

Oxidation: The process of incorporating oxygen into a molecule, such a substrate of a P450 oxygenase. Specific types of oxidation include, for example, epoxidation and hydroxylation. **"Epoxidation"** involves a chemical reaction in which an oxygen atom is joined to an olefinically unsaturated molecule to form a cyclic, three-membered ether. An "olefin" is a hydrocarbon containing a carbon-carbon double bond. **"Hydroxylation"** is a chemical reaction in which a hydroxyl (-OH) group is incorporated into a molecule.

Oxygenase activity: Enzymes exhibiting oxygenase activity are capable of directly incorporating oxygen into a substrate molecule. The process of incorporating oxygen into a substrate molecule is called **"oxidation."** Oxygenases can be either dioxygenases, in which case the oxygenase incorporates two oxygen atoms into the substrate; or, monooxygenases, in which only one oxygen atom is incorporated into the primary substrate, for example, to form a hydroxyl or epoxide group. Monooxygenases also may be referred to as **"hydroxylases."** Taxoid oxygenases are a subset of oxygenases that specifically utilize taxoids as substrates. Taxoid oxygenases can utilize, for example, taxoid substrates having a methylene group at any position, including for example, taxoids having a 5-methylene group (such as, taxoid 5α-hydroxylases), taxoids having a 2-methylene group (such as, taxoid 2α-hydroxylases), taxoids having a 7-methylene group (such as, taxoid 7β-hydroxylases), taxoids having a 10-methylene group (such as, taxoid 10β-hydroxylasess), taxoids having a 13-methylene group (such as, taxoid 13α-hydroxylases), or taxoids having a 14-methylene group (such as, taxoid 14β-hydroxylases).

Oxygenases: Oxygenases are enzymes that display oxygenase activity as describe above. A particular oxygenase may recognize one or more substrates. An oxygenase that will recognize more than one substrate is said to have "relaxed substrate specificity." Different oxygenases may recognize the same substrate and have "shared substrate specificity." Oxygenase enzyme activity assays may utilize one or more different substrates depending on the specificity(ies) of the particular oxygenase enzyme. One of ordinary skill in the art will appreciate that a variety of general oxygenase activity assays, including, for instance, the spectrophotometry-based assay described herein, are available, and that direct assays can be used to test oxygenase catalysis directed towards different substrates.

Polypeptide: A polymer in which the monomers are amino acid residues joined together through amide bonds. When the amino acids are alpha-amino acids, either the L-optical isomer or the D-optical isomer can be used, the L-isomers being preferred. The term polypeptide or protein as used herein encompasses any amino acid sequence and includes modified sequences such as glycoproteins. The term polypeptide is specifically intended to cover naturally occurring proteins, as well as those that are recombinantly or synthetically produced. The term(s) **"isolated polypeptide"** (or isolated protein) as used herein refers to a polypeptide that is substantially free of other proteins, lipids, carbohydrates or other materials with which it is naturally associated. In one embodiment, the

polypeptide is at least 50%, for example at least 70%, at least 80%, at least 90%, or at least 95%, free of other proteins, lipids, carbohydrates or other materials with which it is naturally associated.

Probes and primers: Nucleic acid probes and primers can be readily prepared based on the nucleic acid molecules provided in this disclosure. A probe comprises a detectable isolated nucleic acid. In some instances, a probe is directly attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, enzyme substrates, co-factors, ligands, chemiluminescent or fluorescent agents, haptens, and enzymes. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed, *e.g.*, in Sambrook *et al.* (In: *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, 1989) and Ausubel *et al.* (In: *Current Protocols in Molecular Biology*, Greene Publ. Assoc. and Wiley-Intersciences, 1992).

Primers are short nucleic acid molecules, such as DNA oligonucleotides 10 nucleotides or more in length. Longer DNA oligonucleotides can be about 15, 17, 20, or 23 nucleotides or more in length. Primers can be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, and then the primer extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, *e.g.*, by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods known in the art.

Methods for preparing and using probes and primers are described, for example, in Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, 1989), Ausubel *et al.* (In *Current Protocols in Molecular Biology*, Greene Publ. Assoc. and Wiley-Intersciences, 1998), and Innis *et al.* (*PCR Protocols, A Guide to Methods and Applications*, Academic Press, Inc., San Diego, CA, 1990). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, © 1991, Whitehead Institute for Biomedical Research, Cambridge, MA). One of ordinary skill in the art will appreciate that the specificity of a particular probe or primer increases with its length. Thus, in order to obtain greater specificity, probes and primers can be selected that comprise at least 17, 20, 23, 25, 30, 35, 40, 45, 50 or more consecutive nucleotides.

Protein: A biological molecule expressed by a gene and comprised of amino acids.

Purified: The term **purified** does not require absolute purity; rather, it is intended as a relative term. Thus, for example, a purified protein preparation is one in which the protein referred to is more pure than the protein in its natural environment within a cell. For example, a preparation of an enzyme can be considered as purified if the enzyme content in the preparation represents at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 95% of the total protein content of the preparation.

Recombinant: A nucleic acid that has a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two otherwise separated segments of sequence. This artificial combination can be accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, *e.g.*, by genetic engineering techniques.

"Recombinant" also is used to describe nucleic acid molecules that have been artificially manipulated, but contain the same control sequences and coding regions that are found in the organism from which the gene was isolated.

Sequence identity: The similarity between two nucleic acid sequences or between two amino acid sequences is expressed in terms of the level of sequence identity shared between the sequences. Sequence identity is typically expressed in terms of percentage identity; the higher the percentage, the more similar the two sequences.

Methods for aligning sequences for comparison are well known in the art. Various programs and alignment algorithms are described in: Smith and Waterman, *Adv. Appl. Math.* 2:482, 1981; Needleman and Wunsch, *J. Mol. Biol.* 48:443, 1970; Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444, 1988; Higgins and Sharp, *Gene* 73:237-244, 1988; Higgins and Sharp, *CABIOS* 5:151-153, 1989; Corpet *et al.*, *Nucleic Acids Research* 16:10881-10890, 1988; Huang, *et al.*, *Computer Applications in the Biosciences* 8:155-165, 1992; Pearson *et al.*, *Methods in Molecular Biology* 24:307-331, 1994; Tatiana *et al.*, (1999), *FEMS Microbiol. Lett.*, 174:247-250, 1999. Altschul *et al.* present a detailed consideration of sequence-alignment methods and homology calculations (*J. Mol. Biol.* 215:403-410, 1990).

The National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST™, Altschul *et al.*. *J. Mol. Biol.* 215:403-410, 1990) is available from several sources, including the National Center for Biotechnology Information (NCBI, Bethesda, MD) and on the Internet, for use in connection with the sequence-analysis programs blastp, blastn, blastx, tblastn and tblastx. A description of how to determine sequence identity using this program is available on the internet under the help section for BLAST™.

For comparisons of amino acid sequences of greater than about 30 amino acids, the "Blast 2 sequences" function of the BLAST™ (Blastp) program is employed using the default BLOSUM62 matrix set to default parameters (cost to open a gap [default = 5]; cost to extend a gap [default = 2]; penalty for a mismatch [default = -3]; reward for a match [default = 1]; expectation value (E) [default = 10.0]; word size [default = 3]; number of one-line descriptions (V) [default = 100]; number of alignments to show (B) [default = 100]). When aligning short peptides (fewer than around 30 amino acids), the alignment should be performed using the Blast 2 sequences function, employing the PAM30 matrix set to default parameters (open gap 9, extension gap 1 penalties). Proteins with even greater similarity to the reference sequences will show increasing percentage identities when assessed by this method, such as at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, or at least 95% sequence identity.

For comparisons of nucleic acid sequences, the "Blast 2 sequences" function of the BLAST™ (Blastn) program is employed using the default BLOSUM62 matrix set to default parameters (cost to open a gap [default = 11]; cost to extend a gap [default = 1]; expectation value (E) [default = 10.0]; word size [default = 11]; number of one-line descriptions (V) [default = 100]; number of alignments to show (B) [default = 100]). Nucleic acid sequences with even greater similarity to the reference sequences will show increasing percentage identities when assessed by this

method, such as at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% sequence identity.

An alternative indication that two nucleic acid molecules are closely related is that the two molecules hybridize to each other under stringent conditions (see "Hybridization" above).

Nucleic acid sequences that do not show a high degree of identity can nevertheless encode similar amino acid sequences, due to the degeneracy of the genetic code. It is understood that changes in nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid molecules that all encode substantially the same protein.

Specific binding agent: An agent that binds substantially only to a defined target. For example, a taxoid 5α-hydroxylase protein-specific binding agent binds substantially only the taxoid 5α-hydroxylase protein.

Antibodies are exemplar specific binding agents. Antibodies can be produced using standard procedures described in a number of texts, including Harlow and Lane (*Antibodies, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, 1988). Shorter fragments of antibodies can also serve as specific binding agents, including, for instance, Fabs, Fvs, and single-chain Fvs (SCFvs). Antibody fragments are defined as follows: (1) Fab, the fragment which contains a monovalent antigen-binding fragment of an antibody molecule produced by digestion of whole antibody with the enzyme papain to yield an intact light chain and a portion of one heavy chain; (2) Fab', the fragment of an antibody molecule obtained by treating whole antibody with pepsin, followed by reduction, to yield an intact light chain and a portion of the heavy chain; two Fab' fragments are obtained per antibody molecule; (3) (Fab')2, the fragment of the antibody obtained by treating whole antibody with the enzyme pepsin without subsequent reduction; (4) F(ab')2, a dimer of two Fab' fragments held together by two disulfide bonds; (5) Fv, a genetically engineered fragment containing the variable region of the light chain and the variable region of the heavy chain expressed as two chains; and (6) single chain antibody (SCA), a genetically engineered molecule containing the variable region of the light chain, the variable region of the heavy chain, linked by a suitable polypeptide linker as a genetically fused single chain molecule. Methods of making these fragments are routine.

Substrate: A molecule that binds to an enzyme, such as a P450 oxygenase, and undergoes a chemical change, such as oxidation, during the ensuing enzymatic reaction. Exemplar substrates for the disclosed P450 oxygenases are described throughout this specification. An "exogenous substrate" is a substrate that is added to a particular type of cell.

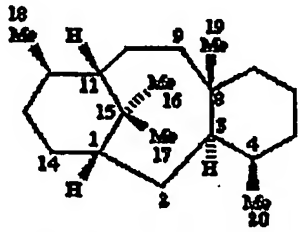
Taxadien-5-ol transacylase activity: Capable of transferring an acyl group (such as an acetyl group) from an acyl carrier (such as acetyl-CoA) to a taxoid substrate comprising a hydroxyl group at C5 (such as taxadien-5α-ol); for additional details, see, *e.g.*, U.S. Pat. No. 6,287,835.

Taxadien-2-ol transacylase activity: Capable of transferring an acyl group (such as a benzoyl group) from an acyl carrier (such as benzoyl CoA) to a taxoid substrate comprising a

hydroxyl group at C2 (such as 2-debenzoyl-7,13-diacetylbaccatin III); for additional details, see, *e.g.*, PCT Pub. No. WO 01/23586.

Taxadiene synthase activity: Capable of cyclizing geranylgeranyl diphosphate to produce taxadiene, as described in detail, *e.g.*, in U.S. Pat. Nos. 6,610,527; 6,114,160; and 5,994,114.

Taxoid: A chemical based on the taxane ring structure (pentamethyl[9.3.1.0]^{3,8}tricyclopentadecane). The core taxane ring structure is described, for example, in Kingston *et al.*, *Progress in the Chemistry of Organic Natural Products*, Springer-Verlag, 1993, and has the chemical structure:



Exemplary taxoids are described throughout the specification and also include, without limitation, taxadiene, taxadienyl acetate (including, *e.g.*, taxa-5α-yl acetate), taxa-4(5),11(12)-diene, taxa-4(20),11(12)-diene, taxadien-5α-ol, taxa-4(20),11(12)-dien-5,13-diol, 5α-acetoxy-10β, 14β-dihydroxy taxadiene, 2-debenzoyl taxane, 10-deacetyl baccatin III, baccatin III, 3'-N-debenzoyltaxol, taxa-4(20),11(12)-dien-5α,9α,10β-triol, taxa-4(20),11(12)-dien-2α,5α-diol (and diacetate ester); taxa-4(20),11(12)-dien-5α,9α,10β,13α-tetraol and corresponding tetraacetate (taxusin tetraol and taxusin, respectively), taxa-4(20),11(12)-dien-5α,9α-diol (and monoacetate and diacetate); taxa-4(20),11(12)-dien-5α,10β-diol (and monoacetate and diacetate); taxa-4(20),11(12)-dien-5α,9α,10β-triol (and acetate esters); and a taxoid having a 5-methylene group (R-CH_2-R).

Transfected: A process by which a nucleic acid molecule is introduced into cell, for instance by molecular biology techniques, resulting in a **transfected (or transformed) cell.** As used herein, the term transfection encompasses all techniques by which a nucleic acid molecule might be introduced into such a cell, including transduction with viral vectors, transfection with plasmid vectors, and introduction of DNA by electroporation, lipofection, and particle gun acceleration.

Vector: A nucleic acid molecule as introduced into a host cell, thereby producing a transfected (or transformed) host cell. A vector can include nucleic acid sequences that permit it to replicate in a host cell, such as an origin of replication. A vector can also include one or more selectable marker genes and other genetic elements known in the art.

Unless otherwise explained, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. The singular terms "a," "an," and "the" include plural referents unless context clearly indicates otherwise. Similarly, the word "or" is intended to include "and" unless the context clearly indicates otherwise. Hence "comprising A or B" means "including A or B," or "including A and B." It is further to be understood that all base sizes or amino acid sizes, and all molecular weight or molecular

mass values, given for nucleic acids or polypeptides are approximate, and are provided for description. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including explanations of terms, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

II. Description of Several Specific Embodiments

Disclosed herein are isolated proteins having taxoid oxygenase activity (such as, taxadiene hydroxylation activity) and the nucleic acid sequences encoding such proteins (including, for example, SEQ ID NO: 1). In some embodiments, the protein comprising an amino acid sequence having at least 80% or at least 95% sequence identity to SEQ ID NO: 2 or comprises the sequence in SEQ ID NO: 2.

Isolated nucleic acid molecules that (i) hybridize under high (or very high) stringency conditions with a nucleic acid probe comprising at least 600 base pairs of SEQ ID NO: 1 and (ii) encode a protein having taxoid oxygenase activity are also contemplated by this disclosure; as are the taxoid oxygenase proteins encoded by such nucleic acid molecules.

Further provide herein are isolated nucleic acid molecules having a sequence at least 80% identical to the nucleic acid sequence in SEQ ID NO: 1 and encoding a protein having taxoid oxygenase activity, such as taxoid 5α-hydroxylase activity. A protein encoded by such a nucleic acid molecule is also disclosed.

Also provided are recombinant nucleic acid molecules, which include a promoter sequence operably linked to a nucleic acid molecule encoding a disclosed taxoid oxygenase protein (such as, a taxoid 5α-hydroxylase). In certain examples, a cell (such as, a plant cell (including a *Taxus* cell or cell line), an insect cell, a bacterium, or a yeast cell) or a non-human transgenic organism (such as a plant, including a plant from the genus *Taxus*) are transformed with the recombinant nucleic acid. In particular examples, the cell is an isolated cell, such as a cell line.

This disclosure includes method of identifying a nucleic acid sequence that encodes a taxoid oxygenase, which involve (i) hybridizing a probe to a nucleic acid sequence under high (or very high) stringency conditions, wherein the probe comprises at least 600 contiguous nucleotides of SEQ ID NO: 1; and (ii) determining that a protein encoded by the nucleic acid sequence is capable of oxidizing a taxoid substrate. A protein capable of oxidizing a taxoid substrate is thereby identified as a taxoid oxygenase. In some examples, oxidizing the taxoid substrate involves hydroxylating the taxoid substrate.

Methods of hydroxylating a substrate are also disclosed. Such methods involve contacting a substrate with at least one oxygenase having an amino acid sequence at least 95% identical to SEQ ID NO: 2 (or having the sequence of SEQ ID NO: 2); and allowing the oxygenase to oxidize the substrate. In some methods, oxidation of the substrate involves hydroxylation of the substrate. In

other methods, the substrate is a taxoid (such as, paclitaxel, a paclitaxel intermediate, a taxadiene, taxa-4(5),11(12)-diene or taxa-4(20),11(12)-diene). In some cases, hydroxylation occurs at position C5 of the taxoid. In specific embodiments, the oxygenase is expressed in an isolated cell or in a transgenic plant, bacterium, insect, fungus or yeast, and the hydroxylation of the substrate occurs *in vivo*. In other embodiments, the substrate is an exogenous substrate, which is fed to the isolated cell, transgenic plant, transgenic bacterium, transgenic insect, transgenic fungus or transgenic yeast.

Also provided herein are methods for increasing paclitaxel yield in a cell (such as a *Taxus* cell, including, for example, a *Taxus* cell line), which involve introducing any of the taxoid oxygenase-encoding recombinant nucleic acid molecules disclosed herein into a paclitaxel-producing cell, wherein the production of paclitaxel is increased in the cell following the introduction of the recombinant nucleic acid molecule. In a particular example, the recombinant nucleic acid molecule that is introduced into the cell includes the nucleic acid sequence in SEQ ID NO: 1. In some examples of this method, the amount of paclitaxel produced by the cell is at least four fold higher following introduction of the recombinant nucleic acid molecule into the cell. In more specific examples, methods for increasing paclitaxel yield in a cell (such as a *Taxus* cell, including, for example, a *Taxus* cell line) further involve introducing additional nucleic acid molecules into the cell. Exemplar additional nucleic acid molecules include those: (i) encoding a protein having taxadiene synthase activity (*e.g.*, nucleic acid molecules having at least 90% sequence identity to SEQ ID NO: 19 (or its protein-coding region), and encoding a protein having taxadiene synthase activity); (ii) encoding a protein having taxadien-5-ol transacylase activity (*e.g.*, nucleic acid molecules having at least 90% sequence identity to SEQ ID NO: 21 (or its protein-coding region), and encoding a protein having taxadien-5-ol transacylase activity); (iii) encoding a protein having taxadien-2-ol transacylase activity (*e.g.*, nucleic acid molecules having at least 90% sequence identity to SEQ ID NO: 23 (or its protein-coding region), and encoding a protein having taxadien-2-ol transacylase activity); (iv) encoding a protein having taxoid oxygenase activity (such as, taxoid 7β-hydroxylase activity, taxoid 14β-hydroxylase activity, taxoid 10β-hydroxylase activity, taxoid 13α-hydroxylase activity, or taxoid 2α-hydroxylase activity) (*e.g.*, nucleic acid molecules having at least 90% sequence identity to any one of the sequences (or their respective protein-coding regions) set forth in SEQ ID NOs: 3, 5, 7, 9, 11, 13, 15, 17, or 40 and encoding a protein having taxoid oxygenase activity (such as, taxoid 7β-hydroxylase activity (*e.g.*, SEQ ID NO: 7), taxoid 14β-hydroxylase activity (*e.g.*, SEQ ID NO: 11), taxoid 10β-hydroxylase activity (*e.g.*, SEQ ID NO: 15), or taxoid 13α-hydroxylase activity (*e.g.*, SEQ ID NO: 17), or taxoid 2α-hydroxylase activity (*e.g.*, SEQ ID NO: 40)); or (v) combinations of (i), (ii), (iii), or (iv). In specific methods, the additional nucleic acid molecules comprise one or more of the nucleic acid sequences set forth in SEQ ID NOs: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 40 (or any combination thereof).

This disclosure also provides antibodies or antibody fragments that bind any of the taxoid oxygenase proteins, such as a taxoid 5α-hydroxylase, described herein. In specific examples the antibody is a monoclonal antibody. In other examples, the antibody fragment is a Fab, F(ab)2, or Fv fragment, or a combination thereof.

III. Taxoid 5α-Hydroxylase Nucleic Acids and Proteins

This disclosure provides P450 oxygenases, such as a taxoid 5α-hydroxylase, and variants thereof, and nucleic acid molecules encoding these proteins, including cDNA sequences.

A nucleic acid molecule encoding a taxoid 5α-hydroxylase and the corresponding deduced amino acid sequence of taxoid 5α-hydroxylase, are shown in SEQ ID NOs: 1 and 2, respectively. The nucleic acid molecule encodes a protein of 502 amino acids in length (SEQ ID NO: 2).

With the provision herein of the sequence of the taxoid 5α-hydroxylase protein (SEQ ID NO: 2) and cDNA (SEQ ID NO: 1), *in vitro* nucleic acid amplification (such as polymerase chain reaction (PCR)) may be utilized as a simple method for producing taxoid 5α-hydroxylase encoding sequences. The following provides representative techniques for preparing cDNA in this manner.

RNA (such as mRNA or total RNA) is extracted from cells by any one of a variety of methods well known to those of ordinary skill in the art. Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, 1989) and Ausubel *et al.* (In *Current Protocols in Molecular Biology*, Greene Publ. Assoc. and Wiley-Intersciences, 1992) provide descriptions of methods for RNA isolation. Taxoid 5α-hydroxylase is expressed, at least, in cells from plants of the genus *Taxus*. Thus, in some examples, RNA may be extracted from *Taxus* cells. The extracted RNA is then used, for example, as a template for performing reverse transcription (RT)-PCR amplification of cDNA. Representative methods and conditions for RT-PCR are described in Kawasaki *et al.*, (In *PCR Protocols, A Guide to Methods and Applications*, Innis *et al.* (eds.), 21-27, Academic Press, Inc., San Diego, California, 1990).

The selection of amplification primers will be made according to the portion(s) of the cDNA that is to be amplified. In one embodiment, primers may be chosen to amplify a segment of a cDNA or, in another embodiment, the entire cDNA molecule. Variations in amplification conditions may be required to accommodate primers and amplicons of differing lengths and composition; such considerations are well known in the art and are discussed for instance in Innis *et al.* (*PCR Protocols, A Guide to Methods and Applications*, Academic Press, Inc., San Diego, CA, 1990). By way of example, the coding portion of the taxoid 5α-hydroxylase cDNA molecule (approximately 1509 base pairs) may be amplified using the following combination of primers:

5'-ATGGACGCCCTGTATAAGAG-3' (forward) (SEQ ID NO: 32)

5'-TCAATTGACTATGGTCTCGG-3' (reverse) (SEQ ID NO: 33)

These primers are illustrative only; one skilled in the art will appreciate that many different primers may be derived from the provided cDNA sequence in order to amplify particular regions of taxoid 5α-hydroxylase cDNA, as well as the complete sequence of the taxoid 5α-hydroxylase cDNA.

Re-sequencing of PCR products obtained by amplification procedures optionally can be performed to facilitate confirmation of the amplified sequence and provide information about natural variation of this sequence in different populations or species. Oligonucleotides derived from the provided taxoid 5α-hydroxylase sequences may be used in such sequencing methods.

Orthologs of the disclosed P450 oxygenases, such as a taxoid 5α-hydroxylase, are likely present in a number of other members of the *Taxus* genus (such as, *T. brevifolia*, *T. canadensis*, *T. baccata*, *T. globosa*, *T. floridana*, *T. wallichiana*, *T. media* and *T. chinensis*) and other taxoid-producing organisms (such as, *Taxomyces andreanae*). With the provision of the disclosed oxygenase nucleic acid sequence, the cloning by standard methods of cDNAs and genes that encode oxygenase orthologs in these other organisms is now enabled. Orthologs of the disclosed oxygenase genes have oxygenase biological activity, including for example oxidation (such as, hydroxylation or epoxidation) of the C5 position of a taxoid. Orthologs will generally share at least 65% sequence identity with the disclosed P450 oxygenase cDNA (for example, SEQ ID NO: 1). Sequence identity will generally be greater in *Taxus* species more closely related to *Taxus cuspidata*. In specific embodiments, orthologous oxygenase (for example, taxoid 5α-hydroxylase) molecules may share at least 70%, at least 75%, at least 80% at least 85%, at least 90%, at least 91%, at least 93%, at least 95%, or at least 98% sequence identity with the disclosed *Taxus cuspidata* nucleotide or amino acid sequences.

Both conventional hybridization and PCR amplification procedures may be utilized to clone sequences encoding oxygenase orthologs. Common to both of these techniques is the hybridization of probes or primers that are derived from the oxygenase nucleic acid sequences. Furthermore, the hybridization may occur in the context of Northern blots, Southern blots, or PCR.

Direct PCR amplification may be performed on cDNA or genomic libraries prepared from the plant species in question, or RT-PCR may be performed using mRNA extracted from the plant cells using standard methods. PCR primers will comprise at least 10 consecutive nucleotides of the oxygenase sequences. One of skill in the art will appreciate that sequence differences between the oxygenase nucleic acid sequence and the target nucleic acid to be amplified may result in lower amplification efficiencies. To compensate for this, longer PCR primers or lower annealing temperatures may be used during the amplification cycle. Whenever lower annealing temperatures are used, sequential rounds of amplification using nested primer pairs may be necessary to enhance specificity.

For conventional hybridization techniques the hybridization probe is preferably conjugated with a detectable label such as a radioactive label, and the probe is preferably at least 10 nucleotides in length. As is well known in the art, increasing the length of hybridization probes tends to give enhanced specificity. The labeled probe derived from the oxygenase nucleic acid sequence may be hybridized to a plant cDNA or genomic library and the hybridization signal detected using methods known in the art. The hybridizing colony or plaque (depending on the type of library used) is purified and the cloned sequence contained in that colony or plaque isolated and characterized.

Orthologs of the oxygenases alternatively may be obtained by immunoscreening of an expression library. With the provision herein of the disclosed oxygenase nucleic acid sequences, the enzymes may be expressed and purified in a heterologous expression system (*e.g.*, *E. coli*) and used to raise antibodies (monoclonal or polyclonal) specific for oxygenases. Antibodies also may be raised against synthetic peptides derived from the oxygenase amino acid sequence presented herein.

Methods of raising antibodies are well known in the art and are described generally in Harlow and Lane, *Antibodies, A Laboratory Manual*, Cold Springs Harbor, 1988. Such antibodies can be used to screen an expression cDNA library produced from a plant. This screening will identify the oxygenase ortholog. The selected cDNAs can be confirmed by sequencing and enzyme activity assays.

Oligonucleotides derived from the taxoid 5α-hydroxylase cDNA sequence (*e.g.*, SEQ ID NO: 1), or fragments of this cDNA, are encompassed within the scope of the present disclosure. Such oligonucleotides may be used, for example, as probes or primers. In one embodiment, oligonucleotides may comprise a sequence of at least 10 consecutive nucleotides of the taxoid 5α-hydroxylase nucleic acid sequence. If these oligonucleotides are used with an *in vitro* amplification procedure (such as PCR), lengthening the oligonucleotides may enhance amplification specificity. Thus, in other embodiments, oligonucleotide primers comprising at least 15, 20, 25, 30, 35, 40, 45, 50, or more consecutive nucleotides of these sequences may be used.

One of ordinary skill in the art will appreciate that the specificity of a particular probe or primer increases with its length. Thus, for example, a primer comprising 30 consecutive nucleotides of a *Taxus cuspidata* taxoid 5α-hydroxylase encoding nucleotide will anneal to a target sequence, such as a taxoid 5α-hydroxylase gene homolog present in a cDNA library from another *Taxus* species (or other paclitaxel-producing species), with a higher specificity than a corresponding primer of only 15 nucleotides. Thus, in order to obtain greater specificity, probes and primers can be selected that comprise at least 17, 20, 23, 25, 30, 35, 40, 45, 50 or more consecutive nucleotides of taxoid 5α-hydroxylase nucleotide sequences. In particular examples, probes or primers can be at least 100, 250, 500, or 600 consecutive nucleic acids of a disclosed 5α-hydroxylase sequence.

Oligonucleotides (such as, primers or probes) may be obtained from any region of a disclosed 5α-hydroxylase nucleic acid sequence. By way of example, the taxoid 5α-hydroxylase cDNA, ORF and gene sequences may be apportioned into about halves, thirds or quarters based on sequence length, and the isolated nucleic acid molecules (*e.g.*, oligonucleotides) may be derived from the first or second halves of the molecules, from any of the three thirds, or from any of the four quarters. The cDNA also could be divided into smaller regions, *e.g.* about eighths, sixteenths, twentieths, fiftieths and so forth, with similar effect. The taxoid 5α-hydroxylase cDNA shown in SEQ ID NO: 1 can be used to illustrate this. The taxoid 5α-hydroxylase cDNA is 1509 nucleotides in length and so in one specific embodiment, it may be hypothetically divided into about halves (nucleotides 40-794 and 795-1548), in another specific embodiment, in about thirds (nucleotides 40-543, 544-1047, and 1048-1548) or, in yet another specific embodiment, in about quarters (nucleotides 40-417, 418-795, 795-1173 and 1174-1548). Alternatively, it may be divided into regions that encode for conserved domains such as, for example, the commonly occurring PERF motif and the region surrounding the invariant, heme-binding cysteine residue (von Wachenfeldt and Johnson, "Structures of eukaryotic cytochrome P450 enzymes," In: *Cytochrome P450: Structure,*

Mechanism, and Biochemistry, 2nd Ed., P.R. Ortiz de Montellano, ed., New York: Plenum, pp. 183-223, 1995).

IV. Cloning of the Taxoid 5α-Hydroxylase Gene

The taxoid 5α-hydroxylase cDNA sequence and fragments described above do not contain introns, upstream transcriptional promoter or regulatory regions or downstream transcriptional regulatory regions of the taxoid 5α-hydroxylase gene. The taxoid 5α-hydroxylase gene may be isolated by routine procedures. For instance, the taxoid 5α-hydroxylase gene may be isolated by homology screening using the cDNA sequence and the BLAST program. Direct sequencing, using the "long-distance sequence method," of one or more BAC or PAC clones that contain the taxoid 5α-hydroxylase sequence can be employed.

Using the information disclosed herein, the regulatory elements flanking the taxoid 5α-hydroxylase gene can be identified and characterized. These regulatory elements may be characterized by standard techniques. In one embodiment, deletion analysis is performed wherein successive nucleotides of a putative regulatory region are removed and the effect of the deletions is studied by transient expression analysis. In another embodiment, the effect of the deletions is studied by long-term expression analysis. The identification and characterization of regulatory elements flanking the genomic taxoid 5α-hydroxylase gene may be made by functional analysis (deletion analyses, *etc.*) in *Taxus* cells by either transient or long-term expression analyses.

It will be apparent to one skilled in the art that either the genomic clone or the cDNA or sequences derived from these clones may be utilized in applications, including but not limited to, studies of the expression of the taxoid 5α-hydroxylase gene, studies of the function of the taxoid 5α-hydroxylase protein, and the generation of antibodies to the taxoid 5α-hydroxylase protein. Descriptions of applications describing the use of taxoid 5α-hydroxylase cDNA, or fragments thereof, are therefore intended to comprehend the use of the genomic taxoid 5α-hydroxylase gene.

It will also be apparent to one of ordinary skill in the art that taxoid 5α-hydroxylase genes may now be cloned from other *Taxus* species by standard cloning methods. In one embodiment, such orthologous taxoid 5α-hydroxylase genes will share at least 65% sequence identity with the taxoid 5α-hydroxylase nucleic acid disclosed herein; and in other embodiments, more closely related orthologous sequences will share at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, or at least 98% sequence identity with this sequence.

V. Taxoid 5α-Hydroxylase Sequence Variants

With the provision of taxoid 5α-hydroxylase protein and corresponding nucleic acid sequences herein, the creation of variants of these sequences is now enabled. Variant oxygenases include proteins that differ in amino acid sequence from the oxygenase sequences disclosed, but that retain oxygenase biological activity.

In one embodiment, variant taxoid 5α-hydroxylase proteins include proteins that differ in amino acid sequence from the taxoid 5α-hydroxylase sequences disclosed but that share at least 70%

amino acid sequence identity with the provided taxoid 5α-hydroxylase protein. In other embodiments, other variants will share at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% amino acid sequence identity. Manipulation of the disclosed taxoid 5α-hydroxylase nucleotide sequence using standard procedures, including in one specific, non-limiting, embodiment, site-directed mutagenesis or in another specific, non-limiting, embodiment, PCR, can be used to produce such variants. The simplest modifications involve the substitution of one or more amino acids for amino acids having similar biochemical properties. These so-called conservative substitutions are likely to have minimal impact on the activity of the resultant protein. The following table shows exemplar conservative amino acid substitutions:

Original Residue	Conservative Substitutions
ala	Ser
arg	Lys
asn	Gln; his
asp	Glu
cys	Ser
gln	Asn
glu	Asp
gly	Pro
his	Asn; gln
ile	Leu; val
leu	ile; val
lys	Arg; gln; glu
met	Leu; ile
phe	Met; leu; tyr
ser	Thr
thr	Ser
trp	Tyr
tyr	Trp; phe
val	ile; leu

In some embodiments, the functional identity of a 5α-hydroxylase variant can be maintained if amino acid substitutions are introduced in regions outside of the conserved domains of the protein, where amino acid substitutions are less likely to affect protein function. FIG. 6 shows the alignment of nine taxoid oxygenase amino acid sequences, including the 5α-hydroxylase sequence disclosed herein. Shaded amino acid residues are conserved among all of the illustrated sequences. In certain embodiments, oxygenase variants share the highly conserved (*i.e.*, shaded and marked by asterisk) amino acid residues shown in FIG. 6. FIG. 6 also demonstrates conservative amino acid variations (*i.e.*, marked by ":") among these taxoid oxygenase sequences. In other embodiments, oxygenase variants having conservative substitutions (as described in the foregoing table) at the amino acid positions indicated by ":" in FIG. 6 are contemplated herein. Amino acid residues that are not highly conserved (*i.e.*, shaded or marked by asterisk in FIG. 6) or conservative variations (*i.e.*, marked by ":" in FIG. 6) are least likely to be functionally relevant and, therefore, may tolerate less conservative amino acid substitutions with little to no effect on the function of the resultant variant. In other embodiments, 5α-hydroxylase protein variants may be designed (as discussed above) based on highly

conserved and conservative amino acids shown in the alignment of three of the foregoing nine amino acid sequences, as shown in FIG. 2.

In another embodiment, more substantial changes in 5α-hydroxylase function or other protein features may be obtained by selecting amino acid substitutions that are less conservative than conservative substitutions. In one specific, non-limiting, embodiment, such changes include changing residues that differ more significantly in their effect on maintaining polypeptide backbone structure (*e.g.*, sheet or helical conformation) near the substitution, charge or hydrophobicity of the molecule at the target site, or bulk of a specific side chain. The following specific, non-limiting, examples are generally expected to produce the greatest changes in protein properties: (a) a hydrophilic residue (*e.g.*, seryl or threonyl) is substituted for (or by) a hydrophobic residue (*e.g.*, leucyl, isoleucyl, phenylalanyl, valyl or alanyl); (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain (*e.g.*, lysyl, arginyl, or histadyl) is substituted for (or by) an electronegative residue (*e.g.*, glutamyl or aspartyl); or (d) a residue having a bulky side chain (*e.g.*, phenylalanine) is substituted for (or by) one lacking a side chain (*e.g.*, glycine).

Variant taxoid 5α-hydroxylase encoding sequences may be produced by standard DNA mutagenesis techniques. In one specific, non-limiting, embodiment, M13 primer mutagenesis is performed. Details of these techniques are provided in Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, 1989), Ch. 15. By the use of such techniques, variants may be created that differ in minor ways from the taxoid 5α-hydroxylase sequences disclosed. In one embodiment, DNA molecules and nucleotide sequences that are derivatives of those specifically disclosed herein, and which differ from those disclosed by the deletion, addition, or substitution of nucleotides while still encoding a protein that has at least 65% sequence identity with the taxoid 5α-hydroxylase encoding sequence disclosed (SEQ ID NO: 1), are comprehended by this disclosure. In other embodiments, more closely related nucleic acid molecules that share at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% nucleotide sequence identity with the disclosed taxoid 5α-hydroxylase sequences are comprehended by this disclosure. Alternatively, related nucleic acid molecules can have no more than 3, 5, 10, 20, 50, 75, or 100 nucleic acid changes compared to SEQ ID NO: 1. In one embodiment, such variants may differ from the disclosed sequences by alteration of the coding region to fit the codon usage bias of the particular organism into which the molecule is to be introduced.

In other embodiments, the coding region may be altered by taking advantage of the degeneracy of the genetic code to alter the coding sequence such that, while the nucleotide sequence is substantially altered, it nevertheless encodes a protein having an amino acid sequence substantially similar to the disclosed taxoid 5α-hydroxylase protein sequences. For example, because of the degeneracy of the genetic code, four nucleotide codon triplets - (GCT, GCG, GCC and GCA) - code for alanine. The coding sequence of any specific alanine residue within the taxoid 5α-hydroxylase protein, therefore, could be changed to any of these alternative codons without affecting the amino acid composition or characteristics of the encoded protein. Based upon the degeneracy of the genetic

code, variant DNA molecules may be derived from the cDNA and gene sequences disclosed herein using standard DNA mutagenesis techniques, as described above, or by synthesis of DNA sequences. Thus, this disclosure also encompasses nucleic acid sequences that encode a taxoid 5α-hydroxylase protein, but which vary from the disclosed nucleic acid sequences by virtue of the degeneracy of the genetic code.

In one embodiment, variants of the taxoid 5α-hydroxylase protein may also be defined in terms of their sequence identity with the prototype taxoid 5α-hydroxylase protein. As described above, taxoid 5α-hydroxylase proteins share at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% amino acid sequence identity with the taxoid 5α-hydroxylase protein (SEQ ID NO: 2). Alternatively, variants of the taxoid 5α-hydroxylase protein can have no more than 3, 5, 10, 15, 20, 25, 30, 40, or 50 amino acid changes compared to SEQ ID NO: 2. Nucleic acid sequences that encode such proteins/fragments readily may be determined simply by applying the genetic code to the amino acid sequence of a taxoid 5α-hydroxylase protein or fragment, and such nucleic acid molecules may readily be produced by assembling oligonucleotides corresponding to portions of the sequence.

Nucleic acid molecules that are derived from the taxoid 5α-hydroxylase cDNA nucleic acid sequences include molecules that hybridize under low stringency, high stringency, or very high stringency conditions to the disclosed prototypical taxoid 5α-hydroxylase nucleic acid molecules, and fragments thereof.

Taxoid 5α-hydroxylase nucleic acid encoding molecules (including the cDNA shown in SEQ ID NO: 1, and nucleic acids comprising this sequence), and orthologs and homologs of these sequences, may be incorporated into transformation or expression vectors.

VI. Introduction of Oxygenases into Plants or Plant Cells

A nucleic acid molecule (such as a cDNA or gene) encoding taxoid 5α-hydroxylase may be incorporated into any organism (intact plant, animal, microbe, *etc.*) or cell or tissue culture system (such as, suspension cell culture, callus cell culture, or immobilized cell culture) for any useful purpose known to those of ordinary skill in the art, including, without limitation, (i) production of taxoid 5α-hydroxylase, (ii) synthesis of 5α-hydroxylated taxoids, such as taxadien-5α-ol; (iii) enhancement of the rate of production and/or the absolute amount of one or more taxoids derived from 5α-hydroxylated taxoids, such as taxadien-5α-ol; (iv) enhancement of the rate of production and/or the absolute amount of paclitaxel or paclitaxel intermediates or derivatives.

In one embodiment, a disclosed 5α-hydroxylase nucleic acid molecule is introduced into a plant or plant cell, for example, a gymnosperm species (such as, a *Taxus* species). Gymnosperms are a useful expression system, at least, because of (i) compatible codon usage for high translational efficiency; (ii) recognition of the encoded preprotein by the plastid import system; (iii) high fidelity in proteolytic processing by the plastids to the mature enzyme form; and (iv) efficient protein-protein

interaction with upstream and downstream enzymes of the paclitaxel pathway for most efficient channeling of metabolites.

After a cDNA (or gene) encoding a protein involved in the determination of a particular plant characteristic has been isolated, standard techniques may be used to express the cDNA in transgenic plants in order to modify the particular plant characteristic. The basic approach is to clone the cDNA into an expression vector, such that the cDNA is operably linked to control sequences (*e.g.*, a promoter), which direct expression of the cDNA in plant cells. The transformation vector is introduced into plant cells by any of various techniques (*e.g.*, electroporation), and progeny plants containing the introduced cDNA are selected. Preferably all or part of the transformation vector stably integrates into the genome of the plant cell. That part of the transformation vector that integrates into the plant cell and that contains the introduced cDNA and associated sequences for controlling expression (the introduced "transgene") may be referred to as the recombinant expression cassette.

Selection of progeny plants containing the introduced transgene may be made based upon the detection of an altered phenotype. Such a phenotype may result directly from the cDNA cloned into the transformation vector or may be manifest as enhanced resistance to a chemical agent (such as an antibiotic) as a result of the inclusion of a dominant selectable marker gene incorporated into the transformation vector.

Successful examples of the modification of plant characteristics by transformation with cloned cDNA sequences are replete in the technical and scientific literature. Selected examples that serve to illustrate the knowledge in this field of technology include, without limitation, U.S. Patent No. 4,459,355 ("Method for Transforming Plant Cells"); U.S. Patent No. 5,571,706 ("Plant Virus Resistance Gene and Methods"); U.S. Patent No. 5,677,175 ("Plant Pathogen Induced Proteins"); U.S. Patent No. 5,510,471 ("Chimeric Gene for the Transformation of Plants"); U.S. Patent No. 5,750,386 ("Pathogen-Resistant Transgenic Plants"); U.S. Patent No. 5,597,945 ("Plants Genetically Enhanced for Disease Resistance"); U.S. Patent No. 5,589,615 ("Process for the Production of Transgenic Plants with Increased Nutritional Value Via the Expression of Modified 2S Storage Albumins"); U.S. Patent No. 5,750,871 ("Transformation and Foreign Gene Expression in Brassica Species"); U.S. Patent No. 5,268,526 ("Overexpression of Phytochrome in Transgenic Plants"); U.S. Patent No. 5,262,316 ("Genetically Transformed Pepper Plants and Methods for their Production"); U.S. Patent No. 5,569,831 ("Transgenic Tomato Plants with Altered Polygalacturonase Isoforms"); U.S. Patent No. 5,932,782 ("Plant Transformation Method Using Agrobacterium Species Adhered to Microprojectiles"); and U.S. Patent No. 6,759,573 ("Method to Enhance Agrobacterium-Mediated Transformation of Plants").

These examples include descriptions of transformation vector selection, transformation techniques, and the construction of constructs designed to over-express the introduced cDNA. In light of the foregoing and the provision herein of the oxygenase amino acid sequences and nucleic acid sequences, it is thus apparent that one of ordinary skill in the art will be able to introduce the cDNAs, or homologous or derivative forms of these molecules, into plants in order to produce plants

having enhanced oxygenase activity. Furthermore, the expression of one or more oxygenases in plants may give rise to plants having increased production of paclitaxel and related compounds.

A. Vector Construction, Choice of Promoters

A number of recombinant vectors suitable for stable transfection of plant cells or for the establishment of transgenic plants have been described, including those described in Weissbach and Weissbach, *Methods for Plant Molecular Biology*, Academic Press, 1989; and Gelvin *et al.*, *Plant and Molecular Biology Manual*, Kluwer Academic Publishers, 1990. Typically, plant transformation vectors include one or more cloned plant genes (or cDNAs) under the transcriptional control of 5'- and 3'-regulatory sequences and a dominant selectable marker. Such plant transformation vectors typically also contain a promoter regulatory region (*e.g.*, a regulatory region controlling inducible or constitutive, environmentally or developmentally regulated, or cell- or tissue-specific expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

Examples of constitutive plant promoters that may be useful for expressing the cDNA include: the cauliflower mosaic virus (CaMV) 35S promoter, which confers constitutive, high-level expression in most plant tissues (see, *e.g.*, Odel *et al.*, *Nature*, 313:810, 1985; Dekeyser *et al.*, *Plant Cell*, 2:591, 1990; Terada and Shimamoto, *Mol. Gen. Genet.*, 220:389, 1990; Benfey and Chua, *Science*, 250:959-966, 1990); the nopaline synthase promoter (An *et al.*, *Plant Physiol.*, 88:547, 1988); and the octopine synthase promoter (Fromm *et al.*, *Plant Cell*, 1:977, 1989).

Agrobacterium-mediated transformation of *Taxus* species has been accomplished, and the resulting callus cultures have been shown to produce paclitaxel (Han *et al.*, *Plant Science*, 95:187-196, 1994). Therefore, it is likely that incorporation of one or more of the described oxygenases under the influence of a strong promoter (like CaMV promoter) would increase production yields of paclitaxel and related taxoids in such transformed cells.

A variety of plant gene promoters that are regulated in response to environmental, hormonal, chemical, and/or developmental signals also can be used for expression of the cDNA in plant cells, including promoters regulated by: (a) heat (Callis *et al.*, *Plant Physiol.*, 88:965, 1988; Ainley, *et al.*, *Plant Mol. Biol.*, 22:13-23, 1993; and Gilmartin *et al.*, *Plant Cell*, 4:839-949, 1992); (b) light (*e.g.*, the pea rbcS-3A promoter, Kuhlemeier *et al.*, *Plant Cell*, 1:471, 1989, and the maize rbcS promoter, Schaffner and Sheen, *Plant Cell*, 3:997, 1991); (c) hormones, such as abscisic acid (Marcotte *et al.*, *Plant Cell*, 1:969, 1989); (d) wounding (*e.g.*, wunI, Siebertz *et al.*, *Plant Cell*, 1:961, 1989); and (e) chemicals such as methyl jasmonate or salicylic acid (see also Gatz *et al.*, *Ann. Rev. Plant Physiol. Plant Mol. Biol.*, 48:9-108, 1997).

Alternatively, tissue-specific (root, leaf, flower, and seed, for example) promoters (Carpenter *et al.*, *Plant Cell*, 4:557-571, 1992; Denis *et al.*, *Plant Physiol.*, 101:1295-1304, 1993; Opperman *et al.*, *Science*, 263:221-223, 1993; Stockhause *et al.*, *Plant Cell*, 9:479-489, 1997; Roshal *et al.*, *EMBO J.*, 6:1155, 1987; Schernthaner *et al.*, *EMBO J.*, 7:1249, 1988; and Bustos *et al.*, *Plant Cell*, 1:839, 1989) can be fused to the coding sequence to obtain a particular expression in respective organs.

Alternatively, the native oxygenase gene promoters may be utilized. With the provision herein of the oxygenase nucleic acid sequences, one of skill in the art will appreciate that standard molecular biology techniques can be used to determine the corresponding promoter sequences. One of skill in the art also will appreciate that less than the entire promoter sequence may be used in order to obtain effective promoter activity. The determination of whether a particular region of this sequence confers effective promoter activity may be ascertained readily by operably linking the selected sequence region to an oxygenase cDNA (in conjunction with suitable 3' regulatory region, such as the NOS 3' regulatory region as discussed below) and determining whether the oxygenase is expressed.

Plant transformation vectors also may include RNA processing signals, for example, introns, that may be positioned upstream or downstream of the ORF sequence in the transgene. In addition, the expression vectors also may include additional regulatory sequences from the 3'-untranslated region of plant genes, *e.g.*, a 3'-terminator region, to increase mRNA stability of the mRNA, such as the PI-II terminator region of potato or the octopine or nopaline synthase (NOS) 3'-terminator regions. The native oxygenase gene 3'-regulatory sequence also may be employed.

As noted above, plant transformation vectors also may include dominant selectable marker genes to allow for the ready selection of transformants. Such genes include those encoding antibiotic resistance genes (*e.g.*, resistance to hygromycin, kanamycin, bleomycin, G418, streptomycin, or spectinomycin) and herbicide resistance genes (*e.g.*, phosphinothricin acetyloxygenase).

B. Arrangement of Taxol Oxygenase Sequence in a Vector

The particular arrangement of the oxygenase sequence in the transformation vector is selected according to the type of expression of the sequence that is desired. In most instances, enhanced oxygenase activity is desired, and the oxygenase ORF is operably linked to a constitutive high-level promoter such as the CaMV 35S promoter. As noted above, enhanced oxygenase activity also may be achieved by introducing into a plant a transformation vector containing a variant form of the oxygenase cDNA or gene, for example a form that varies from the exact nucleotide sequence of the oxygenase ORF, but that encodes a protein retaining an oxygenase biological activity.

C. Transformation and Regeneration Techniques

Transformation and regeneration of a wide variety of plant species, including gymnosperms, angiosperms, monocots and dicots are now routine (see, *e.g.*, Glick and Thompson, eds., *Methods in Plant Molecular Biology*, CRC Press, Boca Raton, Fla., 1993), and the appropriate transformation technique can be determined by the practitioner. The choice of method varies with the type of plant to be transformed; those skilled in the art will recognize the suitability of particular methods for given plant types. Suitable methods may include, but are not limited to: electroporation of plant protoplasts (*e.g.*, Rhodes *et al.*, *Science*, 240(4849):204-207, 1988); liposome-mediated transformation; polyethylene glycol (PEG)-mediated transformation (*e.g.*, Lyznik *et al.*, *Plant Mol. Biol.*, 13:151-161, 1989); transformation using viruses (*e.g.*, Brisson *et al.*, *Nature*, 310:511-514, 1984); microinjection of plant cells (*e.g.*, de la Pena *et al.*, *Nature*, 325:274-276, 1987); micro-projectile bombardment of plant cells (Klein *et al.*, *Plant Physiol.*, 91:440-444, 1989; Boynton *et al.*, *Science*, 240(4858):1534-

1538,1988); vacuum infiltration; and *Agrobacterium tumefaciens* (AT)-mediated transformation. Exemplar procedures for transforming and regenerating plants are described, for instance, in the patent documents listed at the beginning of this section. Additionally, plant transformation strategies and techniques are reviewed by Birch (*Ann. Rev. Plant Phys. Plant Mol. Biol.*, 48:297, 1997), and Forester *et al.* (*Exp. Agric.*, 33:15-33, 1997).

In particular embodiments, transformation of *Taxus* species can be achieved, for example, by employing the methods of Han *et al.* (*Plant Science*, 95:187-196, 1994).

D. Selection of Transformed Plants

Following transformation and regeneration of plants with the transformation vector, transformed plants or cells can be selected using a selectable marker incorporated into the transformation vector. In some examples, such a marker confers antibiotic resistance on the seedlings of transformed plants, and selection of transformants can be accomplished by exposing the seedlings to appropriate concentrations of the antibiotic. For instance, a commonly used selectable marker gene is neomycin phosphotransferase II (NPT II), which confers resistance to the antibiotic, kanamycin. Another selectable marker gene which can be employed is the gene which confers resistance to the herbicide glufosinate (Basta). A screenable gene commonly used is the 0-glucuronidase gene (GUS). The presence of this gene is characterized using a histochemical reaction in which a sample of putatively transformed cells is treated with a GUS assay solution. After an appropriate incubation, the cells containing the transformation vector (which includes the GUS gene) turn blue.

After transformed plants are selected and grown to maturity, they can be assayed using the methods described herein to assess production levels of paclitaxel and other taxoids.

VII. Production of Recombinant Taxoid Oxygenase in Heterologous Expression Systems

Various commonly known systems are available for heterologous expression of the disclosed 5α-hydroxylase nucleic acid molecules to yield the encoded proteins, including, eukaryotic and prokaryotic expression systems. In some examples, eukaryotic expression systems are used to facilitate posttranslational modification of the expressed protein and/or to direct the expressed protein to a desired cellular compartment.

Methods of expressing proteins in heterologous expression systems are well known in the art. Typically, a nucleic acid molecule encoding all or part of the protein of interest, such as a 5α-hydroxylase, is obtained using methods such as those described herein. The protein-encoding nucleic acid sequence is cloned into an expression vector that is suitable for the particular host cell of interest using standard recombinant DNA procedures. Expression vectors include (among other elements) regulatory sequences (*e.g.*, promoters) that can be operably linked to the desired protein-encoding nucleic acid molecule to cause the expression of such nucleic acid molecule in the host cell. Together, the regulatory sequences and the protein-encoding nucleic acid sequence are an "expression cassette." Expression vectors may also include an origin of replication, marker genes

that provide phenotypic selection in transformed cells, one or more other promoters, and a polylinker region containing several restriction sites for insertion of heterologous nucleic acid sequences.

Expression vectors useful for expression of heterologous protein(s) in a multitude of host cells are well known in the art, and some specific examples are provided herein. The host cell is transfected with (or infected with a virus containing) the expression vector using any method suitable for the particular host cell. Such transfection methods are also well known in the art and non-limiting exemplar methods are described herein. The transfected (also called, transformed) host cell is capable of expressing the protein encoded by the corresponding nucleic acid sequence in the expression cassette. Transient or stable transfection of the host cell with one or more expression vectors is contemplated by the present disclosure.

The cloned expression vector encoding one or more of the disclosed oxygenases may be transformed into any of various cell types for expression of the cloned nucleotide. Many different types of cells may be used to express modified nucleic acid molecules. Examples include cells of yeasts, fungi, insects, mammals, and plants, including primary cells and immortal cell lines. For instance, common mammalian cells that could be used include HeLa cells, SW-527 cells (ATCC deposit #7940), WISH cells (ATCC deposit #CCL-25), Daudi cells (ATCC deposit #CCL-213), Mandin-Darby bovine kidney cells (ATCC deposit #CCL-22) and Chinese hamster ovary (CHO) cells (ATCC deposit #CRL-2092). Common yeast cells include *Pichia pastoris* (ATCC deposit #201178) and *Saccharomyces cerevisiae* (ATCC deposit #46024). Insect cells include cells from *Drosophila melanogaster* (ATCC deposit #CRL-10191), the cotton bollworm (ATCC deposit #CRL-9281), and *Trichoplusia ni* egg cell homoflagellates. Fish cells that may be used include those from rainbow trout (ATCC deposit #CLL-55), salmon (ATCC deposit #CRL-1681), and zebrafish (ATCC deposit #CRL-2147). Amphibian cells that may be used include those of the bullfrog, *Rana catesbelana* (ATCC deposit #CLL-41). Reptile cells that may be used include those from Russell's viper (ATCC deposit #CCL-140). Plant cells that could be used include *Chlamydomonas* cells (ATCC deposit #30485), *Arabidopsis* cells (ATCC deposit #54069), tomato plant cells (ATCC deposit #54003) and *Taxus* cells (including, *e.g.*, cells from *T. cuspidata*, *T. brevifolia*, *T. canadensis*, *T. baccata*, *T. globosa*, *T. fioridana*, *T. wallichiana*, *T. media* and *T. chinensis*). Many of these cell types are commonly used and are available from the ATCC as well as from commercial suppliers such as Pharmacia (Uppsala, Sweden), and Invitrogen.

Expressed protein may be accumulated within a cell or may be secreted from the cell. Such expressed protein may then be collected and purified. This protein may be characterized for activity and stability and may be used to practice any of the various methods disclosed herein. Further details of some specific embodiments are discussed below.

A. Yeast

Various yeast strains and yeast-derived vectors are used commonly for the expression of heterologous proteins. For instance, *Pichia pastoris* expression systems, obtained from Invitrogen (Carlsbad, California), may be used to express the disclosed P450 oxygenases, such as a taxoid 5α-hydroxylase. Such systems include suitable *Pichia pastoris* strains, vectors, reagents,

transformants, sequencing primers, and media. Available strains include KM71H (a prototrophic strain), SMD1168H (a prototrophic strain), and SMD1168 (a pep4 mutant strain) (Invitrogen Product Catalogue, 1998, Invitrogen, Carlsbad CA).

Saccharomyces cerevisiae, is another yeast that is commonly used in heterologous expression systems. The plasmid YRp7 (Stinchcomb *et al.*, *Nature*, 282:39,1979; Kingsman *et al.*, *Gene*, 7:141, 1979; Tschemper *et al.*, *Gene*, 10:157, 1980) is commonly used as an expression vector in *Saccharomyces*. This plasmid contains the trp1 gene that provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, such as strains ATCC No. 44,076 and PEP4-1 (Jones, *Genetics*, 85:12, 1977). The presence of the trp1 lesion as a characteristic of the yeast host cell genome then provides an effective environment for detecting transformation by growth in the absence of tryptophan.

Yeast host cells can be transformed using the polyethylene glycol method, as described by Hinnen (*Proc. Natl. Acad. Sci. USA*, 75:1929, 1978). Additional yeast transformation protocols are set forth in Gietz *et al.* (*Nucl. Acids Res.*, 20(17):1425, 1992) and Reeves *et al.* (*FEMS*, 99(2-3):193-197, 1992).

Suitable promoting sequences in yeast vectors include the promoters for 3-phosphoglycerate kinase (Hitzeman *et al.*, *J. Biol. Chem.*, 255:2073, 1980) or other glycolytic enzymes (Hess *et al.*, *J. Adv. Enzyme Reg.*, 7:149, 1968; Holland *et al.*, *Biochemistry*, 17:4900, 1978), such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase. In the construction of suitable expression vectors, the termination sequences associated with these genes are also ligated into the expression vector 3' of the sequence desired to be expressed to provide polyadenylation of the mRNA and termination. Other promoters that have the additional advantage of transcription controlled by growth conditions are the promoter region for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and the aforementioned glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Any plasmid vector containing yeast-compatible promoter, origin of replication and termination sequences is suitable.

Non-yeast eukaryotic vectors may be used with equal facility for expression of proteins encoded by modified nucleotides according to the invention. Mammalian vector/host cell systems containing genetic and cellular control elements capable of carrying out transcription, translation, and post-translational modification are well known in the art. Examples of such systems are the well known baculovirus system, the ecdysone-inducible expression system that uses regulatory elements from *Drosophila melanogaster* to allow control of gene expression, and the sindbis viral expression system that allows high-level expression in a variety of mammalian cell lines, all of which are available from Invitrogen (Carlsbad, California).

B. Baculovirus-Infected Insect Cells

Another representative eukaryotic expression system involves the recombinant baculovirus, *Autographa californica* nuclear polyhedrosis virus (AcNPV; Summers and Smith, *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures*, 1986; Luckow *et al.*, *Biotechnol.*, 6:47-55, 1987). Infection of insect cells (such as cells of the species *Spodoptera frugiperda*) with the recombinant baculoviruses results in the expression taxoid 5α-hydroxylase protein in the insect cells. Baculoviruses do not infect humans and can therefore be safely handled in large quantities.

A baculovirus expression vector is prepared as previously described using standard molecular biology techniques. The vector may comprise the polyhedron gene promoter region of a baculovirus, the baculovirus flanking sequences necessary for proper crossover during recombination (the flanking sequences comprise about 200-300 base pairs adjacent to the promoter sequence) and a bacterial origin of replication which permits the construct to replicate in bacteria. In particular examples, the vector is constructed so that (i) the taxoid 5α-hydroxylase protein-encoding nucleic acid sequence is operably linked to the polyhedron gene promoter (collectively, the "expression cassette") and (ii) the expression cassette is flanked by the above-described baculovirus flanking sequences.

Insect host cells (such as, *Spodoptera frugiperda* cells) are infected with a recombinant baculovirus and cultured under conditions allowing expression of the baculovirus-encoded taxoid 5α-hydroxylase. The expressed oxygenase may, if desired, be extracted from the insect cells using methods known in the art.

C. Mammalian Cells

Mammalian host cells may also be used for heterologous expression of a disclosed oxygenase, such as a taxoid 5α-hydroxylase. Examples of suitable mammalian cell lines include, without limitation, monkey kidney CVI line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line 293S (Graham *et al.*, *J. Gen. Virol.*, 36:59, 1977); baby hamster kidney cells (BHK, ATCC CCL 10); Chinese hamster ovary cells (Urlab and Chasin, *Proc. Natl. Acad. Sci USA*, 77:4216, 1980); mouse sertoli cells (TM4, Mather, *Biol. Reprod.*, 23:243, 1980); monkey kidney cells (CVI-76, ATCC CCL 70); African green monkey kidney cells (VERO-76, ATCC CRL-1587); human cervical carcinoma cells (HELA, ATCC CCL 2); canine kidney cells (MDCK, ATCC CCL 34); buffalo rat liver cells (BRL 3A, ATCC CRL 1442); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); mouse mammary tumor cells (MMT 060562, ATCC CCL 5 1); rat hepatoma cells (HTC, MI.54, Baumann *et al.*, *J. Cell Biol.*, 85:1, 1980); and TRI cells (Mather *et al.*, *Annals N.Y. Acad. Sci.*, 383:44, 1982). Expression vectors for these cells ordinarily include (if necessary) DNA sequences for an origin of replication, a promoter located in front of the gene to be expressed, a ribosome binding site, an RNA splice site, a polyadenylation site, and/or a transcription terminator site.

Promoters used in mammalian expression vectors can be of viral origin. Such viral promoters may be derived from polyoma virus, adenovirus 2, and simian virus 40 (SV40). The SV40 virus contains two promoters that are termed the early and late promoters. These promoters are useful because they are both easily obtained from the virus as one nucleic acid fragment that also contains the viral origin of replication (Fiers *et al.*, *Nature*, 273:113, 1978). Smaller or larger SV40 DNA fragments may also be used, provided they contain the approximately 250-bp sequence extending from the HindIII site toward the BglI site located in the viral origin of replication. Alternatively, promoters that are naturally associated with the foreign gene (homologous promoters) may be used provided that they are compatible with the host cell line selected for transformation.

An origin of replication may be obtained from an exogenous source, such as SV40 or other virus (*e.g.*, polyoma virus, adenovirus, VSV, BPV) and inserted into the expression vector. Alternatively, the origin of replication may be provided by the host cell chromosomal replication mechanism.

D. Prokaryotes

Prokaryotes may also be used as host cells. Prokaryotic expression systems are useful for (among other things) rapid production of large amounts of plasmid DNA, for production of single-stranded DNA templates used for site-directed mutagenesis, for screening many mutants simultaneously, and for DNA sequencing of the mutants generated. Suitable prokaryotic host cells include, without limitation, *E. coli* K12 strain 94 (ATCC No. 31,446), *E. coli* strain W3 110 (ATCC No. 27,325), *E. coli* X1776 (ATCC No. 31,537), and *E. coli* B; however many other strains of *E. coli*, such as HB101, JM101, NM522, NM538, NM539, and many other species and genera of prokaryotes including bacilli such as *Bacillus subtilis*, other enterobacteriaceae, such as *Salmonella typhimurium* or *Serratia marcesans*, and various *Pseudomonas* species may all be used as hosts.

Prokaryotic host cells or other host cells with rigid cell walls may be transformed using any method known in the art, including, for example, calcium phosphate precipitation, or electroporation. Representative prokaryote transformation techniques are described in Dower (*Genetic Engineering, Principles and Methods*, 12:275-296, Plenum Publishing Corp., 1990) and Hanahan *et al.* (*Meth. Enzymol.*, 204:63, 1991).

Plasmids typically used for transformation of *E. coli* include, without limitation, pBR322, pUC18, pUC19, pUCI18, pUC119, Bluescript M13 and derivatives thereof. Numerous such plasmids are commercially available and are well known in the art. Representative promoters used in prokaryotic vectors include the β-lactamase (penicillinase) and lactose promoter systems (Chang *et al.*, *Nature*, 375:615, 1978; Itakura *et al.*, *Science*, 198:1056, 1977; Goeddel *et al.*, *Nature*, 281:544, 1979), a tryptophan (trp) promoter system (Goeddel *et al.*, *Nucl. Acids Res.*, 8:4057, 1980), and the alkaline phosphatase system.

E. Heterologous Protein Trafficking

Trafficking sequences from plants, animals and microbes can be employed to direct the expression of a disclosed oxygenase, such as a 5α-hydroxylase, to the cytoplasm, endoplasmic

reticulum, mitochondria or other cellular compartment, or to target the protein for export to the medium.

Many eukaryotic proteins contain an endogenous signal sequences. The nucleic acid sequence encoding a signal sequence may be obtained as a restriction fragment from any gene encoding a protein with a signal sequence. By ligating DNA encoding a signal sequence to the 5' end of the DNA encoding a protein of interest, the resultant chimeric protein can be directed to the destination conveyed by the signal sequence.

The signal sequences of several eukaryotic genes are known, including, for example, human growth hormone, proinsulin, and proalbumin (see, *e.g.*, Stryer, *Biochemistry*, Third Edition, W.H. Freeman and Company, New York, N.Y., p. 769, 1988), and can be used as signal sequences in appropriate eukaryotic host cells. Yeast signal sequences, such as acid phosphatase (Arima *et al.*, *Nucl. Acids Res.*, 11:1657, 1983), α-factor, alkaline phosphatase and invertase, may be used to direct secretion from yeast host cells. Prokaryotic signal sequences from genes encoding, for example, LamB or OmpF (Wong *et al.*, *Gene*, 68:193, 1988), MalE, PhoA, or β-lactamase, as well as other genes, may be used to target proteins from prokaryotic cells into the culture medium.

VIII. Production of an Antibody to a Taxoid 5α-hydroxylase Protein

Monoclonal or polyclonal antibodies may be produced to either the normal taxoid 5α-hydroxylase protein or variants of this protein. In one embodiment, antibodies raised against the taxoid 5α-hydroxylase protein would specifically detect the taxoid 5α-hydroxylase protein. That is, such antibodies would recognize and bind the taxoid 5α-hydroxylase protein, or fragments thereof, and would not substantially recognize or bind to other proteins found in *Taxus* cells. In some embodiments, antibodies against the *Taxus cuspidata* taxoid 5α-hydroxylase protein may recognize taxoid 5α-hydroxylase from other paclitaxel-producing species (*e.g.*, *Taxomyces andreanae*), and *vice versa*. Antibodies to the disclosed oxygenase enzymes, and fragments thereof, may be also useful for purification of the enzymes.

The determination that an antibody specifically binds to an antigen is made by any one of a number of standard immunoassay methods; for instance, Western blotting (see, Sambrook *et al.* (eds.), *Molecular Cloning: A Laboratory Manual*, 2nd ed., vols. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989). To determine that a given antibody preparation (such as a preparation produced in a mouse against SEQ ID NO: 2) specifically detects the oxygenase by Western blotting, total cellular protein is extracted from cells and electrophoresed on an SDS-polyacrylamide gel. The proteins are electrophoretically transferred to a membrane (for example, nitrocellulose), and the antibody preparation is incubated with the membrane. After washing the membrane to remove non-specifically bound antibodies, the presence of specifically bound antibodies is detected by the use of a detector molecule (such as, an anti-mouse antibody conjugated to an enzyme such as alkaline phosphatase). Antibodies that specifically detect an oxygenase will be

shown, by this technique, to bind substantially only the oxygenase band (having a position on the gel determined by the molecular weight of the oxygenase).

Substantially pure oxygenase suitable for use as an immunogen can be isolated from transfected cells, transformed cells, or from wild-type cells. Concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms per milliliter. Alternatively, peptide fragments of an oxygenase may be utilized as immunogens. Such fragments may be synthesized chemically using standard methods, or may be obtained by cleavage of the whole oxygenase enzyme followed by purification of the desired peptide fragments. Peptides as short as three or four amino acids in length are immunogenic when presented to an immune system in the context of a Major Histocompatibility Complex (MHC) molecule, such as MHC class I or MHC class II. Accordingly, peptides comprising at least 3 and preferably at least 4, 5, 6 or more consecutive amino acids of the disclosed oxygenase amino acid sequences may be employed as immunogens for producing antibodies.

Because naturally occurring epitopes on proteins frequently comprise amino acid residues that are not adjacently arranged in the peptide when the peptide sequence is viewed as a linear molecule, it may be advantageous to utilize longer peptide fragments from the oxygenase amino acid sequences for producing antibodies. Thus, for example, peptides that comprise at least 10, 15, 20, 25, or 30 consecutive amino acid residues of the amino acid sequence may be employed. Monoclonal or polyclonal antibodies to the intact oxygenase, or peptide fragments thereof may be prepared as described below.

A. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibodies can be prepared from murine hybridomas according to the classical method of Kohler and Milstein (*Nature*, 256:495-497, 1975) or derivative methods thereof. In one specific, non-limiting embodiment, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody-producing cells of the spleen isolated. The spleen cells are fused with mouse myeloma cells using polyethylene glycol, and the excess, non-fused, cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). Successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate, where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall (*Enzymol.*, 70(A):419-439, 1980), and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Harlow and Lane (*Antibodies, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, 1988).

B. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogeneous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein (for instance, expressed using a method described herein), which, in one specific, non-limiting embodiment, can be modified to

enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. In one embodiment, small molecules may tend to be less immunogenic than others and may require the use of carriers and adjuvant, examples of which are known. In another embodiment, host animals may vary in response to site of inoculations and dose, with either inadequate or excessive doses of antigen resulting in low titer antisera. In one specific, non-limiting embodiment, a series of small doses (ng level) of antigen administered at multiple intradermal sites may be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis *et al.* (*J. Clin. Endocrinol. Metab.*, 33:988-991, 1971).

In one embodiment, booster injections will be given at regular intervals, and antiserum harvested when antibody titer thereof begins to fall, as determined semi-quantitatively (for example, by double immunodiffusion in agar against known concentrations of the antigen). See, for example, Ouchterlony *et al.* (In *Handbook of Experimental Immunology*, Wier, D. (ed.) chapter 19. Blackwell, 1973). In one specific, non-limiting embodiment the plateau concentration of antibody is usually in the range of about 0.1 to 0.2 mg/ml of serum (about 12 μM). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher (*Manual of Clinical Immunology*, Ch. 42, 1980).

C. Antibodies Raised against Synthetic Peptides

A third approach to raising antibodies against the taxoid 5α-hydroxylase protein is to use synthetic peptides synthesized on a commercially available peptide synthesizer based upon the predicted amino acid sequence of the taxoid 5α-hydroxylase protein. Polyclonal antibodies can be generated by injecting such peptides into, for instance, rabbits.

D. Antibodies Raised by Injection of Taxoid 5α-Hydroxylase Encoding Sequence

In one embodiment, antibodies may be raised against the taxoid 5α-hydroxylase protein by subcutaneous injection of a recombinant DNA vector that expresses the taxoid 5α-hydroxylase protein into laboratory animals, such as mice. In one specific, non-limiting embodiment, delivery of the recombinant vector into the animals may be achieved using a hand-held form of the Biolistic system (Sanford *et al., Particulate Sci. Technol.*, 5:27-37, 1987), as described by Tang *et al.* (*Nature*, 356:152-154, 1992). In other embodiments, expression vectors suitable for this purpose may include those that express the taxoid 5α-hydroxylase encoding sequence under the transcriptional control of either the human β-actin promoter or the cytomegalovirus (CMV) promoter.

IX. Methods of Using 5α-Hydroxylase

The creation of recombinant vectors and transgenic organisms expressing vectors disclosed herein are useful for controlling the production of the disclosed oxygenases, such as the 5α-hydroxylase. These vectors can be used to decrease oxygenase production or to increase oxygenase production. Increased production of oxygenase can be achieved by including at least one additional oxygenase encoding sequence in the vector. These vectors can be introduced into a host cell, thereby altering oxygenase production. In the case of increased production, the resulting

oxygenase may be used in *in vitro* systems, as well as *in vivo* for increased production of paclitaxel, other taxoids, intermediates of the paclitaxel biosynthetic pathway, and other products.

A. Production of Paclitaxel or Other Taxoid *In Vivo*

One attractive alternative to yew harvest and/or paclitaxel semisynthesis is the production of paclitaxel and taxoids *in vivo*, such as in transgenic organisms and/or cell culture (including, for example, *Taxus* cell culture). Cell culture, for example, lends itself to vat fermentation format (potentially as a continuous process), a high level of process control, and ease of product isolation and purification. This practice further provides the possibility of biochemical/molecular manipulation to direct biosynthesis to specific taxoid precursors, modified forms, and derivatives.

In current practice at the small scale, *Taxus* cell cultures produce about 10-100 mg/L of paclitaxel (up to 1 gram total taxoids/L) in production runs of about 7-10 days; however, production levels are quite variable and not sustainable with time or at scale. Commercially viable production levels of paclitaxel are estimated to be between about 200-400 mg/L and of precursors for semi-synthesis in the range of about 400-800 mg/L range. Enhancement of production levels and/or redirection of taxoid metabolism can be useful to achieve economic viability. Preferably, production levels are consistent and reliable. A system that is biochemically manipulable can permit synthesis of a range of taxoid derivatives (*e.g.*, alternative precursors and second generation drugs). Such a system is now enabled by the disclosure of the 5α-hydroxylase protein and nucleic acid sequences. This enzyme is believed to catalyze a slow-step in the paclitaxel biosynthetic pathway; thus, alone and in combination with other enzymes of the paclitaxel pathway, the disclosed 5α-hydroxylase protein and nucleic acid sequences permit molecular genetic manipulation (genetic engineering) of cultured cells, such as *Taxus* cells, to increase yields of paclitaxel and to direct the pathway to desirable taxoid metabolites.

Production of paclitaxel and related taxoids (such as, taxoid-5-ols, including isomers of taxadien-5-ol) *in vivo* can be accomplished by transfecting a host cell, such as one derived from the *Taxus* genus, with a vector capable of expressing, at least, a disclosed oxygenase (such as, a taxoid 5α-hydroxylase). Methods of making and using suitable expression vectors and transforming a variety of cell types with such vectors have been described above. In certain examples, heterologous or homologous oxygenase sequences are placed under the control of a constitutive promoter, or an inducible promoter; thus, any naturally occurring feedback that might otherwise downregulate oxygenase expression under natural conditions will be eliminated.

In some methods, the host cell does not produce any paclitaxel prior to transfection, in which case, particular methods can involve feeding taxoids (such as, paclitaxel intermediates) to the cell. In other methods, a host cell will express a detectable amount of paclitaxel prior to transfection so that transfection with the expression vector increases the production of paclitaxel in the transfected cell. In particular examples of these methods, paclitaxel production in a transfected cell may be increased by at least two fold, such as at least four fold, at least 10 fold, at least 20 fold, at least 50 fold or at least 100 fold.

A host cell, which has been (or will be) transfected with a disclosed 5α-hydroxylase, may also be transfected (using either the same or a different expression vector) with nucleic acid sequences encoding other enzymes having activities useful for the biosynthesis of paclitaxel including, for example, taxadiene synthase (such as SEQ ID NO: 19 or the protein-encoding portion thereof), taxadienol acetyl transferase (such as, SEQ ID NO: 21 or the protein-encoding portion thereof; also known as TAX1 or TAT), 2-debenzoyl-7,13-diacetylbaccatin III-2-O-benzoyl transferase (such as, SEQ ID NO: 23 or the protein-encoding portion thereof; also known as TAX2), 10-deacetylbaccatin III-10-O-acetyl transferase (such as, SEQ ID NO: 34 or the protein-encoding portion thereof; also known as TAX6 or DBAT), taxoid 13-phenylpropanoyltransferase (such as, SEQ ID NO: 36 or the protein-encoding portion thereof; also known as TAX7); 3'-N-debenzoyltaxol N-benzoyltransferase (such as, SEQ ID NO: 38 or the protein-encoding portion thereof; also known as TAX10 or DBNTBT); or any of a several taxoid oxygenases, including, without limitation, taxoid 7β-hydroxylase, taxoid 14β-hydroxylase, taxoid 10β-hydroxylase, taxoid 13α-hydroxylase, or taxoid 2α-hydroxylase (such as, one or more of the taxoid oxygenases set forth in SEQ ID NOs: 3, 5, 7, 9, 11, 13, 15, 17 or 40). Variants of each of the foregoing enzymes, which maintain the function of the prototype enzyme, would be equally suitable for use in the above-described multi-gene expression system. Such variants can be, for instance, at least 70%, at least 80%, at least 90%, or at least 95% percent identical to either the nucleic acid sequence encoding, or the amino acid sequence of, the prototype enzyme.

Methods and constructs for the introduction of multiple protein-encoding nucleic acids sequences (such as, cDNAs) into cells, such as plant cells (including, *e.g.*, *Taxus* cell lines), using single or multiple transformation event(s) have been described (see, *e.g.*, U.S. Pat. No. 6,337,431; U.S. Pat. Pub. No. 20020129400, U.S. Pat. Pub. No. 20020059660; de Felipe, *Curr. Gene Ther.*, 2(3):355-378, 2002). For example, techniques commonly used for introduction of multiple genes into cells include: (i) co-transformation with mixed multiple plasmid vectors containing different protein-encoding sequences using any transfection method known in the art (*e.g.*, Chen *et al.*, *Nat. Biotechnol.*, 16:1060-1064, 1998; Ye *et al.*, *Science*, 287:303-305, 2000); (ii) sequential re-transformation of the same recipient cell (or cell population) with vectors where each vector contains one or a few protein-encoding sequences (*e.g.*, Lapierre *et al.*, *Plant Physiol.*, 119:153-163, 1999); or sexual crossing between transgenic organisms carrying different transgenes to recombine the genes to a single organism (*e.g.*, Ma *et al.*, *Science*, 268:716-719, 1995); and (iii) linking of multiple genes of different sources into the same vector using conventional molecular cloning technology for transformation (*e.g.*, Van Engelen *et al.*, *Plant Mol. Biol.*, 26:1701-1710, 1994; Daniell and Dhingra, *Curr. Opin. Biotechnol.*, 13:136-141 2001). In particular examples, a multi-gene construct includes a promoter, nucleic acid sequences encoding two or more proteins, inteins, and transcription termination sequences and, optionally, sequences encoding targeting sequences, or tissue specific sequences, such as tissue-specific targeting peptides.

Cells may be transfected (*i.e.*, transformed) with one or more constructs useful for the expression of multiple protein-encoding sequences in a single cell in any manner known in the art or

as described herein including, without limitation, *Agrobracterium* transformation of plant cells (see, for instance, Han *et al.*, *Plant Sci.*, 95(2):187-196, 1994).

B. Production of Paclitaxel or Other Taxoids *In Vitro*

Currently, paclitaxel is produced by a semisynthetic method described in Hezari and Croteau, *Planta Medica*, 63:291-295, 1997. This method involves extracting 10-deacetyl-baccatin III, or baccatin III, intermediates in the paclitaxel biosynthetic pathway, and then finishing the production of paclitaxel using chemical techniques. With the provision of a taxoid 5α-hydroxylase herein, it is now possible to utilize this enzyme (and its variants) to hydroxylate taxoids (such as, paclitaxel intermediates) to produce taxoid-5-ols (including for example, taxadien-5-ol isomers). Such taxoid-5-ols can be used, for example, to facilitate the production of paclitaxel and related taxoids.

In vitro methods involve transfection of a host cell with a vector expressing a disclosed 5α-hydroxylase, as described previously. Following transfection, the recombinant enzyme may hydroxylate available taxoid substrates (including, *e.g.*, taxadiene isomers such as, taxa-4(5),11(12)-diene, and taxa-4(20),11(12)-diene). Such substrates can be naturally present in the host cell (such as, a *Taxus* cell) or can be administered to the host cell, for example, by adding the exogenous substrate to the media bathing the cells. Under these circumstances, 5α-hydroxylation of the substrate can occur *in vivo* (as discussed in the preceding section) and the resultant product, such as a taxadien-5-ol, can be extracted and/or purified for further *in vitro* processing, including the synthesis of paclitaxel, paclitaxel intermediates, or other taxoids.

In other methods, the 5α-hydroxylase protein can be isolated from transfected cells. The isolated protein can, then, be used as a reagent in reactions involving the 5α-hydroxylation of taxoid substrates, including taxadiene isomers (such as, taxa-4(5),11(12)-diene, and taxa-4(20),11(12)-diene).

Embodiments of the invention are illustrated by the following non-limiting Examples.

EXAMPLES

Example 1

HOMOLOGY-BASED CLONING OF CYTOCHROME P450 OXYGENASES FROM *TAXUS*

Previous studies have used DD-RT PCR to obtain cytochrome P450 taxoid oxygenase clones from methyl jasmonate-induced *Taxus* cells. The DD-RT PCR method is limited because it may fail to identify transcripts that are not highly induced by the inducing agent used in the method. This example describes a strategy for cloning taxoid oxygenases, such as the disclosed taxoid 5α-hydroxylase, that is not subject to the biases of DD-RT PCR (Udvardi *et al.*, *Plant Physiol.*, 105:755-756, 1994; Holton and Lester, *Methods Enzymol.*, 272:275-283, 1996; Pauli and Kutchan, *Plant J.* 13:793-801, 1998). This strategy is based upon two highly conserved regions of P450 oxygenase proteins, the commonly occurring PERF motif and the region surrounding the invariant, heme-binding cysteine residue (von Wachenfeldt and Johnson, "Structures of eukaryotic cytochrome

P450 enzymes," In: *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 2nd Ed., P.R. Ortiz de Montellano, ed., New York: Plenum, pp. 183-223, 1995).

Unless expressly stated otherwise, enzymes and reagents used in this and other Examples were obtained from Gibco/BRL (Grand Island, NY), Invitrogen (Carlsbad, CA), New England Biolabs (Beverly, MA) and Stratagene (La Jolla, CA), as indicated, and were used according to the respective manufacturer's instructions. Other chemicals were purchased from Merck (Darmstadt, Germany) and Sigma (St. Louis, MO).

Degenerate and inosine-containing oligonucleotide forward primers directed to the PERF motif and its variant forms were designed based on amino acid sequence alignments of cytochrome P450 oxygenases of plant origin, which were available in the public databases. The following forward primers directed to the PERF motif were synthesized:

5'-TTY MGI CCI AGM GIT TYG AR-3' (SEQ ID NO: 25)

5'-TTY MGI CCI TCI MGI TTY GAR-3' (SEQ ID NO: 26)

5'-CKI III CCI GCI CCR AAI GG-3' (SEQ ID NO: 27)

5'-GAR GAR TTY MGN CCN GAR MG-3' (SEQ ID NO: 28)

5'-GAR AAR TTY III CCI GAI ARG TTY (SEQ ID NO: 29)

Using a similar strategy, degenerate and inosine-containing oligonucleotide reverse primers directed to the conserved heme-binding region were designed and synthesized. These reverse primers are:

5'-GGR CAI III CKI III CCI CCI CCR AAI GG-3' (SEQ ID NO: 30)

5'-CCI GGR CAI ATI MKY YTI CCI GCI CCR AAI GG-3' (SEQ ID NO: 31).

Amplification (Pauli and Kutchan, *Plant J.*, 13:793-801, 1998), using first strand cDNA template derived from mRNA isolated from *T. cuspidata* cells 16 hours post-induction with methyl jasmonate (Ketchum *et al.*, *Biotechnol. Bioengin.*, 62:97-105, 1999; Schoendorf *et al.*, *Proc. Natl. Acad. Sci. USA*, 98:1501-1506, 2001), yielded amplicons of the predicted size (*i.e.*, about 200 base pairs). The amplicons were gel purified, ligated into pGEM-T™ (Promega, Madison, WI), and transformed into *E. coli* JM109 cells for plasmid preparation and insert sequencing.

Based on the amplicon sequences, probes of 40 to 50 nucleotides in length were synthesized, 5'-labeled with [^{32}P]dCTP (ICN, Irvine, CA) using T4 polynucleotide kinase (New England Biolabs, Beverly, MA), and used to screen the previously described induced *T. cuspidata* λ-ZAPII™ cDNA library (Schoendorf *et al.*, *Proc. Natl. Acad. Sci. USA*, 98:1501-1506, 2001) by employing Rapid-Hyb™ (Amersham Pharmacia, Piscataway, NJ) solution. Following 3 rounds of screening, 32 positive plaques were *in vivo* excised as pBluescript II SK(–)™ phagemids in accordance with the manufacturer's (Stratagene) protocol, and partially sequenced using T3 and T7 promoter primers. Based on sequence information, clones that were previously obtained by the DD-RT PCR screen (Schoendorf *et al.*, *Proc. Natl. Acad. Sci. USA*, 98:1501-1506, 2001) were set aside and not further examined. These new clones were obtained in full-length form by Marathon™ 5'-RACE (Clontech, Palo Alto, CA), as necessary, and were fully sequenced.

One clone, designated S1, represented the most abundant cytochrome P450 cDNA isolated by this homology-based cloning approach. The S1 clone was 1688 base pairs in length (GenBank accession no. AY289209) and contained an apparent ORF of 1509 base pairs encoding a predicted protein of 502 amino acids with deduced molecular weight of 56,859 Daltons. The deduced amino acid sequence of clone S1 exhibited characteristics commonly known in the art to be typical of cytochrome P450 enzymes (von Wachenfeldt and Johnson, "Structures of eukaryotic cytochrome P450 enzymes," In: *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 2nd Ed., P.R. Ortiz de Montellano, ed., New York: Plenum, pp. 183-223, 1995), including the oxygen-binding domain (amino acid residues 270-285), an N-terminal membrane anchor (amino acid residues 1-30), the highly conserved heme-binding motif (amino acid residues 433-441) with PFG element (amino acids at positions 437-439), and the absolutely conserved cysteine at position 445.

Comparisons of the clone S1 deduced amino acid sequence with the amino acid sequences of previously characterized cytochrome P450 taxoid hydroxylases, including the taxoid 10β-hydroxylase (GenBank Accession No. AF318211), 13α-hydroxylase (GenBank Accession No. AY056019) and 14β-hydroxylase (GenBank Accession No. AY188177), revealed overall identities in the 61-63% range and similarities in the 79-81% range (FIG. 2).

The sequence analyses described in this Example provide strong evidence that clone S1 encodes a taxoid oxygenase.

Example 2

CYTOCHROME P450 cDNA EXPRESSION IN YEAST

This Example demonstrates one method for readily expressing taxoid oxygenases, such as taxoid 5α-hydroxylase, in yeast.

For functional expression in *Saccharomyces cerevisiae*, the deduced ORFs of clone S1 was amplified by PCR using a gene-specific forward primer (containing the ATG start codon) and a corresponding reverse primer in which the stop codon was deleted to permit read-through when transferred to the expression vector, pYES2.1/V5-HIS-TOPO™ (Invitrogen) (see, *e.g.*, SEQ ID NOs: 32 and 33).

The clone S1 ORF amplicon was cloned into pYES2.1/V5-HIS-TOPO™ using standard techniques. Vector sequences in frame with the cloned S1 ORF encode the simian V5 epitope and a histidine (His_6) tag. Thus, the resultant expression vector (referred to herein as pYES2.1/S1-V5-HIS) encodes a fusion protein containing the complete clone S1 protein with a C-terminal simian V5 epitope and histidine (His_6) tag. This tagging procedure allows detection of the expressed enzyme via immunoblot analysis of the microsomal protein preparation using commercially available antibodies, and has been shown not to compromise the activity of other recombinant taxoid hydroxylases (Jennewein *et al.*, *Arch. Biochem. Biophys.*, 413:262-270, 2003). The pYES2.1/S1-V5-HIS insert was sequenced using the Gal1 (forward) and V5 C-term (reverse) primers

(available from Invitrogen) to confirm that expected S1 ORF sequence was present and in the correct orientation for expression.

The verified S1 clone expression vector was transformed into *Saccharomyces cerevisiae* strain WAT11 using the lithium acetate method (Ito *et al.*, *J. Bacteriol.*, 153:163-168, 1983). The WAT11 strain harbors a galactose-inducible NADPH-cytochrome P450 reductase from *Arabidopsis thaliana*, which is required for efficient reductive coupling to the cytochrome (Pompon *et al.*, *Methods Enzymol.*, 272:51-64, 1996). This yeast expression system also permits testing of catalytic activity by *in vivo* feeding of taxoid substrates to the transformed yeast (Schoendorf *et al.*, *Proc. Natl. Acad. Sci. USA*, 98:1501-1506, 2001), thereby eliminating the need for microsome isolation in preliminary functional screening assays (as discussed in more detail in Example 3).

Transformed yeast cells were grown to stationary phase in 2 ml of SGIA medium at 30°C with 250 rpm mixing. The cells were then harvested by centrifugation (2000g, 10 minutes) and the cell pellet was suspended in 3 ml YPLA galactose-containing induction medium. Approximately 9 hours after induction, the cells were harvested again by centrifugation.

For immunoblotting, the cells were resuspended in lysis buffer (100 mM Tris HCl, pH 8.5, containing 1 mM DTT and 10% v/v glycerol), lysed by sonication (VirSonic, microtip probe, medium setting, 3 x 30 sec, VirTis Co., Gardiner, NY) or by use of a Bead Beater (Biospec Products, Bartlesville, OK), and the microsomes prepared (Pompon *et al.*, *Methods Enzymol.*, 272, 51-64, 1996). Protein (50 µg) was then separated by SDS-PAGE (10% denaturing gel), transferred by wet transfer blotting to nitrocellulose and immobilized by UV-crosslinking. The blot was serially incubated with mouse Penta-His-specific antibody (Qiagen, Valencia, CA) as primary antibody, and alkaline phosphatase-conjugated AffiniPure™ goat anti-mouse IgG (Jackson ImmunoResearch, West Grove, PA) as secondary antibody for detection. The Qiagen protocols were used throughout, with His-size markers as reference, and protein preparations from transformed cells harboring empty vector as negative controls.

A single protein of approximately 57 kDa was specifically identified by Western blot. The observed molecular weight of this recombinant protein agrees with the calculated molecular weight of the deduced S1 protein sequence (see Example 1).

Example 3

IN SITU SCREENING FOR CYTOCHROME P450 FUNCTION

This Example demonstrates that clone S1 can be efficiently expressed in yeast and that taxa-4(5),11(12)-diene and taxa-4(20),11(12)-diene are exemplar high affinity substrates for the clone S1 oxygenase.

Following confirmation of clone S1 expression by immunoblot analysis (see Example 2), the activity of the recombinant cytochrome P450 enzyme was demonstrated by *in vivo* feeding as previously described by Schoendorf *et al.* (*Proc. Natl. Acad. Sci. USA*, 98:1501-1506, 2001). This *in vivo* feeding protocol eliminated the uncertainties associated with microsome isolation and *in vitro*

assay, including the instability of P450 oxygenases in yeast membranes (Schoendorf *et al.*, *Proc. Natl. Acad. Sci. USA*, 98:1501-1506, 2001).

Transformed and induced yeast cells were harvested by centrifugation as described in Example 2. As a negative control for the feeding experiments, the yeast host was transformed with the pYES2.1/V5-HIS-TOPO™ vector containing a β-glucuronidase insert instead of the cytochrome P450 S1 clone. The cells were resuspended in 3 ml of fresh YPLA medium to which 30 μM of the labeled test substrate was added. Test substrates included:

Substrate (Activity)	Activity (Ci/mol)	Reference
(±)-[20-^{3}H]taxa-4(5),11(12)-diene	5.3	Rubenstein *et al.*, *J. Label. Compds. Radiopharm.*, 43:481-491, 2000
(±)-[20-^{3}H]taxa-4(20),11(12)-diene	2.6	Rubenstein *et al.*, *J. Label. Compds. Radiopharm.*, 43:481-491, 2000
(±)-taxa-4(20),11(12)-dien-5α-ol	2.0	Rubenstein *et al.*, *J. Label. Compds. Radiopharm.*, 43:481-491, 2000
(±)-taxa-4(20),11(12)-dien-5α-yl acetate	2.0	Walker *et al.*, *Arch. Biochem. Biophys.*, 364:273-279, 1999; Lovy Wheeler *et al.*, *Arch. Biochem. Biophys.*, 390:265-278, 2001
(±)-taxa-4(20),11(12)-dien-5α-acetoxy-10β-ol	2.0	Jennewein *et al.*, *Proc. Natl. Acad. Sci. USA*, 98:13595-13600, 2001; Jennewein *et al.*, *Arch. Biochem. Biophys.*, 413:262-270, 2003
(±)-taxa-4(20),11(12)-dien-5α,13α-diol	2.0	Jennewein *et al.*, *Proc. Natl. Acad. Sci. USA*, 98:13595-13600, 2001; Jennewein *et al.*, *Arch. Biochem. Biophys.*, 413:262-270, 2003
(+)-[^{3}H-acetyl]taxusin[1]	10.0	De Case *et al.*, *Chem. Commun.*, 1282-1294, 1969; Chau *et al.*, *Chem. Biol.*, 11:663-672, 2004; Chau and Croteau, *Arch. Biochem. Biophys.*, 427:48-57, 2004

[1]Tetraacetate of taxa-4(20),11(12)-dien-5α,9α,10β,13α-tetraol

The cell and test substrate suspension was incubated overnight at 30°C with mixing (250 rpm). The incubation mixture was then treated for 15 minutes in a sonication bath and extracted twice with 3 ml of hexane:ethyl acetate (4:1 v/v). The organic extract was then dried under N_2, the residue dissolved in 100 μl of acetonitrile, and an aliquot was separated by reversed-phase radio-HPLC (250 mm x 4.6 mm column of Alltech (Deerfield, IL) Econosil C_{18} (5 μm); flow rate of 1 ml/min; with radio-detection of the effluent (Flow-One-Beta Series A-1000, Radiomatic Corp., Meriden, CT)). The following gradient was employed: 0-5 minutes at 100% Solvent A (97.99% H_2O with 2% CH_3CN and 0.01% H_3PO_4 (v/v)), 5-15 minutes at 0-50% Solvent B (99.99% CH_3CN with 0.01% H_3PO_4 (v/v)), 15-55 minutes at 50-100% Solvent B, 55-65 minutes at 100% Solvent B, 65-70 minutes at 0-100% Solvent A, 70-75 minutes at 100% Solvent A. The HPLC eluant was collected in 1 minute fractions and the appropriate fractions containing the radiolabeled product were combined, dried under a stream of N_2, and dissolved in the minimum volume of benzene for GC-MS analysis.

GC-MS analyses were performed on a Hewlett-Packard 6890 GC-MSD system using a ZB-5 capillary column (Phenomenex (Torrance, CA); 30 m length; 0.25 mm inner diameter; coated with a 0.25 μm film of phenyl (5%) polysiloxane). Cool on-column injection was used, with He flow rate of 0.7 ml/min and a temperature program from 40°C to 320°C at 20°C/min. Spectra were recorded at 70 eV.

Radio-HPLC analysis showed that the two taxadiene isomers were most efficiently (almost quantitatively in the case of the 4(5),11(12)-isomer) converted to more polar products. In the case of taxa-4(5),11(12)-diene as substrate, the principal biosynthetic product (>92%) eluted with a retention time identical to that of taxa-4(20),11(12)-dien-5α-ol (Hefner *et al.*, *Chem. Biol.*, 3:479-489, 1996; Rubenstein *et al.*, *J. Label. Compds. Radiopharm.*, 43:481-491, 2000) and the minor product (<5%) eluted with a retention time consistent with that of a taxadien-diol. GC-MS analysis (electron impact ionization) confirmed the major product to possess a retention time and mass spectrum identical to that of authentic taxa-4(20),11(12)-dien-5α-ol (Hefner *et al.*, *Chem. Biol.*, 3:479-489, 1996) with characteristic ions at m/z 288 (P+), 273 (P+-CH_3), 270 (P+-H_2O) and 255 (P+-H_2O-CH_3). The minor, more polar product yielded a mass spectrum consistent with that of a taxadien-diol (ions corresponding to the loss of a methyl and two molecules of water from an unobserved parent ion of m/z 304).

In the case of taxa-4(20),11(12)-diene as substrate, the major product (~90%) was again shown, upon radio-HPLC analysis, to possess a retention time identical to that of taxa-4(20),11(12)-dien-5α-ol, and this identification was confirmed as before by GC-MS analysis. The taxadien-diol side product was also observed (~8%), as were a range of other minor metabolites (at ~2% of the product mix) that were also derived from this substrate in the negative control (yeast that expressed β-glucuronidase). These negative controls did not produce taxa-4(20),11(12)-dien-5α-ol or the taxadien-diol from either taxadiene isomer.

A. 13α-Hydroxylase Utilizes 5α-Hydroxylase Product

The order of oxygenation reactions on the taxane (taxadiene) nucleus en route to paclitaxel is not precisely known. However, based on comparison of the structures of the several hundred naturally occurring taxanes (Kingston *et al.*, *The Taxane Diterpenoids*, in Herz *et al.* (eds.), *Progress in the Chemistry of Organic Natural Products*, Springer-Verlag, New York, Vol. 61, p. 206, 1993; and Baloglu *et al.*, *J. Nat. Prod.* 62:1448-1472, 1999), it can be deduced from relative abundances of taxoids with oxygen substitution at each position (Floss *et al.*, *Biosynthesis of Taxol*, in Suffness (ed.), *Taxol: Science and Applications*, CRC Press, Boca Raton, FL, pp. 191-208, 1995) that oxygens at C5 (carbon numbers shown in Section I, "Taxoid") and C10 are introduced early, followed by oxygenation at C2, C9 and C13. Oxygenations at C7 and C1 of the taxane nucleus are considered to be very late introductions, possibly occurring after oxetane ring formation; however, epoxidation (at C4/C20) and oxetane formation seemingly must precede oxidation of the C9 hydroxyl to a carbonyl (Floss *et al.*, *Biosynthesis of Taxol*, in Suffness (ed.), *Taxol: Science and Applications*, CRC Press, Boca Raton, FL, pp. 191-208, 1995).

The taxa-4(20),11(12)-dien-5α-ol radiolabeled product of the clone S1 enzyme was isolated by HPLC, and the purified material was fed to yeast that functionally express the previously characterized taxoid 13α-hydroxylase (Jennewein *et al.*, *Proc. Natl. Acad. Sci. USA*, 98:13595-13600, 2001). As discussed above, 13α-hydroxylation of a paclitaxel intermediate is believed to follow 5α-hydroxylation in the paclitaxel biosynthetic pathway (Floss and Mocek, *Taxol: Science and Applications*, CRC Press, Boca Raton, 191-208, 1995; and Croteau *et al.*, *Curr. Top. Plant Physiol.* 15:94-104, 1996). Thus, as expected, taxa-4(20),11(12)-dien-5α-ol was quantitatively converted to taxa-4(20),11(12)-dien-5α,13α-diol by the 13α-hydroxylase.

This Example demonstrates that cytochrome P450 clone S1 encodes a taxoid 5α-hydroxylase, which catalyzes, at least, the first oxygenation step of the paclitaxel biosynthetic pathway.

Example 4

SUBSTRATE BINDING AND KINETIC ANALYSIS OF RECOMBINANT 5α-HYDROXYLASE

This Example demonstrates that the clone S1 hydroxylase binds, at least, taxa-4(20),11(12)-diene and taxa-4(5),11(12)-diene with high affinity, and efficiently catalyzes both taxadiene isomers to the corresponding taxadien-5α-ol.

To prepare sufficient oxygenase enzyme for comparative analysis of substrate binding and kinetic phenomena, in a host less prone to interfering activity and artifact formation, the taxadiene 5α-hydroxylase S1 cDNA clone was transferred to the previously described baculovirus-*Spodoptera fugiperda* (Sf9) expression system (which also coexpresses a *Taxus* cytochrome P450 reductase) (Jennewein *et al.*, *Proc. Natl. Acad. Sci. USA*, 98:13595-13600, 2001).

For construction of the recombinant baculovirus harboring cytochrome P450 clone S1, the S1 ORF was amplified using Pfu DNA polymerase and gene-specific primers containing a BamHI site immediately upstream of the start codon and another containing a NotI site downstream of the stop codon. The gel purified S1 amplicon was subcloned first into the pCR-Blunt™ vector (Invitrogen) and the insert was then excised using the BamHI/NotI restriction sites and ligated into the similarly digested pFastBac1™ vector (Life Technologies, Grand Island, NY). This S1 pFastBac1™ construct was then used to prepare recombinant Bacmid DNA by transforming *Escherichia coli* strain DH10Bac (Life Technologies) carrying the baculovirus genome. As a negative control for this expression system, recombinant baculovirus containing a β-glucuronidase gene, instead of the cytochrome P450 S1 ORF, was used. Baculovirus construction and transfection of Sf9 cells were carried out according to the Life Technologies protocols, and culturing was performed as previously described (Jennewein *et al.*, *Proc. Natl. Acad. Sci. USA*, 98:13595-13600, 2001).

For microsome preparation, Sf9 cells were harvested three days after transfection, washed twice with 50 mM KH_2PO_4, pH 7.5, containing 9% (w/v) NaCl, twice with 50 mM Hepes, pH 7.5,

containing 0.5 mM EDTA, 0.1 mM DTT and 10% (v/v) glycerol, and then lysed by gentle sonication as before in 50 ml of the Hepes buffer system. Cell debris was removed by centrifugation (10,000g, 20 minutes, 4°C), and the resulting supernatant was then centrifuged at 28,000g (20 minutes, 4°C) and then at 105,000g (120 minutes, 4°C) to provide the microsomal membranes which were resuspended in the same Hepes buffer system without EDTA, or other buffer system as noted herein. Protein content was determined by the Bradford method (Bradford, *Anal. Biochem.*, 72:248-254, 1976) using bovine serum albumin as standard.

The CO-difference spectra of microsomes isolated from Sf9 insect cells expressing either the recombinant 5α-hydroxylase or β-glucuronidase gene were obtained as described by Omar and Sato (*J. Biol. Chem.* 239, 2370-2378, 1964) using a Perkin-Elmer Lambda 18 spectrophotometer (Haudenschild *et al.*, *Arch. Biochem. Biophys.*, 379:127-136, 2000). Based on CO-difference spectra, more than 300 pmol cytochrome P450/mg microsomal protein was routinely produced by this Sf9 insect cell system.

Binding spectra for both taxadiene isomers (in the absence of NADPH) were then recorded using the Sf9 cell microsomes enriched in the recombinant 5α-hydroxylase. Substrate binding spectra were obtained as described by Schenkman and Jansson, *Methods Mol. Biol. (Cytochrome P450 Protocols)*, 107:25-33, 1998) using a Perkin-Elmer Lambda 18 spectrophotometer (Haudenschild *et al.*, *Arch. Biochem. Biophys.*, 379:127-136, 2000). Substrate binding spectra were recorded with up to 200 pmol of recombinant microsomal cytochrome P450 enzyme (as determined by CO-difference spectral analysis) per cuvette in 100 mM sodium phosphate buffer at pH 7.5. In preparation for binding studies, the taxadiene isomers were each dissolved in DMSO and 1 μl additions to the sample were made to a final concentration of 1.5% (v/v). For data analysis, Spectrum for Windows (Perkin-Elmer Corp., Wellesey, MA) and Sigmaplot 7.0 (SPSS Inc., Chicago, IL) were employed and experiments were run in triplicate.

Evaluation of the substrate binding constant (Ks) over a 100-fold range of substrate concentrations showed Ks to vary somewhat from 3 to 5 μM for taxa-4(20),11(12)-diene and from 5 to 8 μM for taxa-4(5),11(12)-diene (a typical data set at 200 pmol protein concentration is illustrated in FIG. 3). These results indicate that the taxadiene 5α-hydroxylase active site binds, at least, both positional isomers of the olefin substrate with high affinity.

Kinetic constants for both isomers were next evaluated (at a saturating 200 μM concentration of NADPH plus regenerating system; Shimada and Yamazaki, *Meth. Mol. Biol.*, 107:85-93, 1998). The isolated microsomes were resuspended in 50 mM Hepes, pH 7.5, containing 1 mM DTT and 5% (v/v) glycerol, and the 1 ml reactions (~600 μg protein, 50 μM substrate dissolved in DMSO, and the requisite cofactors (*e.g.*, NADPH plus regenerating system) were run as described previously, with the identical protocols for product analysis (Jennewein *et al.*, *Proc. Natl. Acad. Sci. USA*, 98:13595-13600, 2001). DMSO was without influence on the reaction. For kinetic evaluation, following the establishment of linear reaction conditions in protein concentration and time, the response to substrate concentration was plotted by the Michaelis-Menten method (Sigmaplot 7.0)

using the calibrated radio-HPLC protocol for product determination. Data from three independent experiments were pooled and the line of best fit taken (R^2 >0.99).

Plotting the lines of best fit (R^2 > 0.99) provided a Km value of 16 ± 3.2 μM, with Vrel of 120, for taxa-4(20),11(12)-diene, and a Km value of 24 ± 2.5 μM, with Vrel of 100, for taxa-4(5),11(12)-diene (see FIG. 4); the latter Km value compares to a Km value of ~6 μM determined previously for the 4(5),11(12)-isomer with the native microsome preparations from yew stem tissue (Hefner *et al.*, *Chem. Biol.* 3:479-489, 1996). Comparison of catalytic efficiencies (Vrel/Km) indicates that both taxa-4(20),11(12)-diene and the 4(5),11(12)-isomer are efficiently catalyzed by the recombinant 5α-hydroxylase enzyme.

The taxadien-diol product, which was observed in the intact yeast system fed the taxadiene substrates (see Example 3), was not observed in the baculovirus-*Spodoptera* system. Thus, it is believed that the diol product results from the action of yeast host enzyme(s) upon the taxadienol produced by the recombinant 5α-hydroxylase; this observation was independently verified by feeding control yeast cells the taxadienol product.

This Example and Example 3 demonstrate that the 5α-hydroxylation is a slow step of paclitaxel biosynthesis relative to the downstream oxygenations and acylations. Embodiments of the disclosed oxygenases catalyze 5α-hydroxylation of several taxoids, including, for example, the natural paclitaxel intermediate, taxa-4(5),11(12)-diene. Thus, recombinant expression of the disclosed oxygenase, for example, in *Taxus* plants and cells will increase pathway flux toward paclitaxel to improve production yields of this drug from its natural, and currently the only commercially viable, source.

Example 5

SUBSTRATE UTILIZATION BY *TAXUS* MICROSOMES

Examples 3 and 4 demonstrate that, at least, two taxadiene isomers are functional substrates of the recombinant clone S1 5α-hydroxylase. *Taxus* cell microsomes contain a structurally uncharacterized 5α-hydroxylase activity (Hefner *et al.*, *Chem. Biol.*, 3:479-489, 1996), which had not been tested previously with the 4(20),11(12)-diene isomer (Hefner *et al.*, *Chem. Biol.*, 3:479-489, 1996). This Example demonstrates that a crude *Taxus* microsome preparation converts both taxa-4(5),11(12)-diene and taxa-4(20),11(12)-diene to the corresponding taxidienols.

Preparation of *Taxus* suspension cell microsomes and assays for microsomal 5α-hydroxylase activity were carried out as previously described (Hefner *et al.*, *Chem. Biol.*, 3:479-489, 1996; Lovy Wheeler *et al.*, *Arch. Biochem. Biophys.*, 390:265-278, 2001) with the following modifications: Unelicited *Taxus media hicksii* cells were harvested 14 days after transfer, separated from the media, frozen in liquid N_2 and ground to a fine powder with a mortar and pestle, with extraction and microsome preparation as described by Lovy Wheeler *et al.* (*Arch. Biochem. Biophys.*, 390:265-278, 2001). The previously described radio-HPLC-based assay (Lovy Wheeler *et al.*, *Arch. Biochem. Biophys.*, 390:265-278, 2001) was employed to separate the substrate from

taxadien-5α-ol and polyols derived therefrom which were summed as "total product" for the purpose of rate determination.

Following the confirmation of linear reaction conditions in protein concentration and time, kinetic constants were determined for both [20-^{3}H]taxa-4(5),11(12)-diene and [20-^{3}H]taxa-4(20),11(12)-diene with the optimized assay (Hefner *et al.*, *Chem. Biol.*, 3:479-489, 1996). The radio-HPLC-based assay previously described by Lovy Wheeler *et al.* (*Arch. Biochem. Biophys.*, 390:265-278, 2001) was employed to permit summing of taxadien-polyols derived subsequently from the initially formed taxadienol product generated by this microsomal system that contains all of the downstream cytochrome P450 taxoid oxygenases of the pathway (Lovy Wheeler *et al.*, *Arch. Biochem. Biophys.*, 390:265-278, 2001). Any kinetic isotope effect (KIE) resulting from the C20 deprotonation of [20-^{3}H]taxa-4(5),11(12)-diene was not considered because previous studies with [20-$^2H_3$]taxa-4(5),11(12)-diene (>99 atom % ^{2}H) indicated that hydrogen removal from C20 is not rate limiting in the overall hydroxylation reaction (Hefner *et al.*, *Chem. Biol.*, 3:479-489, 1996).

By this approach, Michaelis-Menten plotting ($R^2 > 0.98$ for the lines of best fit) revealed a Km value of 48 μM, and Vrel of 100, for taxa-4(5),11(12)-diene, and a Km value of 27 μM, with Vrel of 150, for taxa-4(20),11(12)-diene (see FIG. 4). Thus, one or more constituents of the crude *Taxus* microsomal protein preparation catalyze the reaction of taxa-4(5),11(12)-diene and taxa-4(20),11(12)-diene to the corresponding taxadienols.

Native and recombinant *Taxus* taxadiene synthase, which is believed to be the first enzyme in the paclitaxel biosynthetic pathway, each produce principally taxa-4(5),11(12)-diene (94%), with very low level co-production of taxa-4(20),11(12)-diene (4.8%) and verticillene (1.2%), and only trace amounts of taxa-3(4),11(12)-diene (Williams *et al.*, *Arch. Biochem. Biophys.*, 379:137-146, 2000). For this reason, taxa-4(5),11(12)-diene is believed to be the natural substrate of the mediator(s) of the putative next step in the pathway, namely, 5α-hydroxylation. This Example demonstrates that the 5α-hydroxylase activity present in *Taxus* microsomes efficiently utilized two taxadiene substrates with the catalytic efficiency (Vrel/Km) of the presumed unnatural substrate, taxa-4(20),11(12)-diene, being slightly higher than the presumed natural substrate, taxa-4(5),11(12)-diene.

Example 6

IN VIVO SUBSTRATE OF 5α-HYDROXYLASE

This Example demonstrates that the relaxed substrate specificity of clone S1 oxygenase extends to both naturally occurring substrates and non-naturally occurring substrates.

As discussed in preceding Examples, recombinant 5α-hydroxylase enzyme (clone S1) and 5α-hydroxylase microsomal activity have relaxed substrate specificity and, for example, efficiently utilize taxa-4(20),11(12)-diene as a substrate. *Taxus* taxadiene synthase (native and recombinant enzyme, and allelic variants (Accession No. AY364469 and Accession No. AY364470)) produces very low levels of taxa-4(20),11(12)-diene (4.8%) (Williams *et al.*, *Arch. Biochem. Biophys.*, 379:137-146, 2000). Nonetheless, taxa-4(20),11(12)-diene could be a productive intermediate *in vivo*

if *Taxus* cells expressed a taxadiene isomerase that catalyzed the conversion of taxa-4(20),11(12)-diene to taxa-4(20),11(12)-diene.

Recombinant taxadiene synthase isoforms and taxadiene synthase allelic variants were expressed in *E. coli*. The preparation and assay of the recombinant taxadiene synthase isoforms were conducted by established methods using capillary GC-MS conditions designed to separate taxadiene positional isomers (Williams *et al.*, *Arch. Biochem. Biophys.*, 379:137-146, 2000).

The assay for microsomal taxa-4(5),11(12)-diene isomerase activity (and the reverse isomerization) was carried out under standard cytochrome P450 oxygenase conditions but in the absence of NADPH or O_2, or in the presence of inhibitory concentrations of CO, miconazole or clotrimazole (under conditions described previously for which the rate of 5α-hydroxylation is negligible (Hefner *et al.*, *Chem. Biol.*, 3:479-489, 1996)). A number of additional, potential cofactors were also tested, including FAD, $FADH_2$, FMN, $FMNH_2$, NAD+, NADH and NADP+ (all at 2.5 mM), as well as $MgCl_2$ (at 5.0 mM). The possibility of pH-dependent isomerization was tested by incubating each isomer (100 μM) in phosphate buffer (pH 4 to 10) for 12 hours at 31°C, with separation of isomers as described previously (Williams *et al.*, *Arch. Biochem. Biophys.*, 379:137-146, 2000).

No isomerization of taxa-4(5),11(12)-diene to the 4(20),11(12)-diene isomer (or *vice versa*) was observed in *Taxus* cell microsomes (or *Spodoptera* microsomes enriched in the recombinant clone S1 5α-hydroxylase) under standard assay conditions but in the absence of NADPH or in the absence of O_2 (N_2 atmosphere plus and O_2 scavenging system), or in the presence of CO, 100 μM miconazole, or 100 μM clotrimazole (all conditions under which hydroxylation activity is negligible), nor was isomerization observed in boiled controls containing all cofactors and reactants. Similarly, no interconversion of either positional isomer was observed in the presence of magnesium ion, NAD+, NADH or NADP+, or flavin cofactors, at pH values ranging from 4 to 10.

This Example indicates that taxa-4(5),11(12)-diene is not appreciably isomerized to taxa-4(20),11(12)-diene under physiological conditions. The migration of the double bond from the 4(5)- to the 4(20)-position in the process of taxadienol formation may, but need not, be an inherent feature of the cytochrome P450 oxygenase reaction with taxa-4(5),11(12)-diene as substrate.

This Example demonstrates that taxa-4(20),11(12)-diene is an adventitious, yet efficient, substrate for the oxygenase encoded by clone S1.

Example 7

PROPOSED 5α-HYDROXYLASE MECHANISM OF ACTION

Previous efforts to evaluate the 5α-hydroxylation reaction by the *Taxus* microsomal activity, by search for an epoxide intermediate and through the use of [20-$^2H_3$]taxa-4(5),11(2)-diene to examine a KIE on the deprotonation step, did not elucidate a possible mechanism of action. Two possible mechanisms include, for example, (i) a preliminary conversion of the 4(5)-double bond of taxa-4(5),11(12)-diene to the corresponding 4(5)-epoxide, followed by ring opening and elimination

of a proton from the C20 methyl group to yield the allylic alcohol product, or (ii) cytochrome P450-mediated abstraction of hydrogen from the C20 methyl of the substrate to yield the allylic radical to which oxygen is added at C5 (FIG. 5) (Hefner *et al.*, *Chem. Biol.*, 3:479-489, 1996).

Though not bound by any particular mechanism of action, the utilization of the isomeric taxa-4(20),11(12)-diene by the recombinant hydroxylase, with efficiency comparable to that of the putative natural substrate (*i.e.*, taxa-4(5),11(12)-diene), suggests a mechanism involving abstraction of a hydrogen radical from C20 (or C5 in the case of the other isomer), leading to the delocalized allylic radical, followed by oxygen insertion selectively from the 5α-face of this radical intermediate to accomplish the rearrangement. Perhaps the somewhat tighter binding of the 4(20)-isomer is a reflection of the ability of this isomer to more closely mimic the allylic radical intermediate.

Embodiments of this disclosure provide taxoid oxygenase proteins and nucleic acid molecules, and methods of isolating, making, and using these molecules. Specific embodiments relate to taxoid 5-hydroxylase proteins and nucleic acid molecules, including, for example, 5α-hydroxylase proteins and nucleic acid molecules. Further embodiments provide methods for producing paclitaxel, or its intermediates and, in particular, to methods of hydroxylating taxoids at position 5. It will be apparent that the precise details of the compositions and methods described may be varied or modified without departing from the spirit of this disclosure. We claim all such modifications and variations that fall within the scope and spirit of the claims below.